

STIC-Biotech/ChemLib

45827

From: Schnizer, Holly
Sent: Tuesday, July 03, 2001 7:47 AM
To: STIC-Biotech/ChemLib
Subject: sequence search for appl. no. 09/490,291

Please search the commercial and interference databases for the following:

SEQ ID NO: 2, 4, 6, 8, 9 and 11

Thank you.

Holly Schnizer
AU 1653
CM1-10B05
305-3722

RELATED
PCT 00/30086

qBel
MB

Point of Contact:
Beverly Shears
Technical Info. Specialist
CM1 12C14 Tel: 308-496-4964

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:14 ; Search time 86.32 Seconds
(without alignments)
51.971 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377
Sequence: 1 MRSHHHHHHGMSASGDLKN.....AKIGDLNNTSGIRPPAAKLN 74

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq.0601.*
1: /SID8/gcgdata/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	279.5	74.1	76 21	AA10562
2	243.5	64.6	76 21	AA10561
3	221	58.6	84 19	AAW43025
4	213	56.5	84 19	AAW43026
5	205	54.4	84 19	AAW43027
6	199	52.8	42 21	AA10558
7	197	52.3	84 19	AAW43028
8	181	48.0	84 19	AAW43024
9	159	42.2	42 21	AAW10557
10	124	32.9	42 19	AAW43029
11	121	32.1	43 21	AAW08364

Result No.	Score	Query Match Length	ID	Description
12	121	32.1	43 21	AAW08383
13	119.5	31.7	103 22	AAW70797
14	118.5	31.4	93 22	AAW70796
15	115	30.5	43 21	AAW08344
16	114.5	30.4	43 21	AAW08375
17	113	30.0	43 21	AAW08362
18	113	30.0	43 21	AAW08365
19	111.5	29.6	43 21	AAW08386
20	110.4	27.6	43 21	AAW08378
21	101.5	26.9	286 21	AAV56997
22	101.5	26.9	286 22	AAV56978
23	96.5	25.6	491 18	AAW13373
24	96	25.5	175 21	AAV44383
25	95	25.2	373 14	AAW39486
26	94	24.9	168 21	AAV44384
27	94	24.9	169 21	AAV44380
28	94	24.9	175 21	AAV44382
29	92	24.4	145 20	AAW94266
30	92	24.4	159 19	AAW31552
31	92	24.4	493 12	AAW13892
32	91.5	24.3	168 21	AAV44385
33	91.5	24.3	169 21	AAV44381
34	91	24.1	18 21	AAW10560
35	91	24.1	379 10	AAW90508
36	90.5	24.0	380 10	AAW90509
37	90.5	24.0	380 10	AAW90510
38	90.5	24.0	380 10	AAW90511
39	90.5	24.0	380 10	AAW90512
40	90	23.9	337 14	AAW39485
41	90	23.9	345 19	AAW31555
42	90	23.9	359 19	AAW61046
43	89.5	23.7	342 14	AAW39487
44	89.5	23.7	342 14	AAW39489
45	89.5	23.7	342 14	AAW39491

ALIGNMENTS

RESULT 1
ID AAB10562
AAB10562 standard: protein; 76 AA.
AC AAB10562;
XX
DT 15-DEC-2000 (first entry)
XX
DE Copolymer block protein insert L2-B.
XX
KW Coiled coil protein; copolymer blocker; vulnary; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW antibiotic; dressing; abrasion; burn.
XX
OS Unidentified.
XX
PN US6090911-A.
XX
PD 18-JUL-2000.
XX
PF 22-OCT-1997; 97US-0956307.
XX
PR 22-OCT-1997; 97US-0956307.
XX
PA (UYMA-) UNIV MASSACHUSETTS.
XX
PI Petka WA, McGrath KP, Tirrell DA.
XX
DR WPI; 2000-542235/49.
XX
PT Synthetic block copolymer with an antibiotic compound, useful for
PT dressing abrasion, burn or non-puncture wound, comprises two
PT alpha-helical protein blocks, random-coil protein block and linker
PT proteins -

XX WPI: 1998-120029/11.
 PT Production of peptide hetero:dimer(s) - in the fabrication of
 XX self-assembling protein-based structural material(s)
 XX
 XX
 XX Disclosure: Fig 4B; 17pp; English.
 XX
 CC AAM43024-28 represent recognition sequences obtained from the library of
 CC recognition sequences coded for by AAV04943. This DNA sequence is
 CC obtained by reverse translation. The codons were chosen to maximise
 CC expression in *Escherichia coli*, and to introduce useful restriction for
 CC subsequent genetic manipulations. The DNA sequence use a "mixed site"
 CC approach at the first base of the codons for amino acids positions 5 and
 CC 7 of the heptad (see below). The formation of a self-assembling
 CC structural material is mediated by the artificial recognition sequences
 CC (e.g. present structure). A method of producing a self-assembling
 CC protein-based structural material comprises mixing a multiplicity of
 CC 2 different artificial peptide sequences, designed so as to
 CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
 CC alpha helical secondary structure having at least 2 heptad units. The
 CC artificial peptides spontaneously heterodimerise into a self-assembled
 CC protein-based structural material. The method is used to produce
 CC nanoscale structural materials designed for specific functions e.g.
 CC membranes, fibres, absorbants, reactive materials, etc. A specific
 CC application is in the fabrication of selectively permeable membranes for
 CC protective gloves and coated fabrics for fuel handlers and other
 CC hazardous materials. Materials for medical implants, wound healants and
 CC other medical treatments would also be feasible using this technology.
 XX
 XX
 XX Sequence 84 AA:
 XX
 XX

```

QY      16  GDLKNKVAQLKRRVRSLKDKAAELKQEVSRLENEIEDLKAKIGDLNN 62
          |||
Db      2  gdlknkvaglkrrkvrslkdkaaelkqevsrleenedlkaklgdlkn 48
Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query Match 58.6%; Score 221; DB 19; Length 84;
Best Local Similarity 97.9%; Pred. No. 1,7e-16;

RESULT 4
AAW43026 standard; Peptide; 84 AA.
ID      AAW43026
XX
AC      AAW43026;
XX
DT      08-MAY-1998 (first entry)
XX
DE      Artificial recognition sequence 3.
XX
KW      Recognition sequence: self-assembly; protein-based structural material;
KM      spontaneous heterodimerisation; coiled-coil alpha helix;
KV      selectively permeable membrane; coated fabric.
XX
OS      Synthetic.
XX
PN      US5712366-A.
XX
PD      27-JAN-1998.
XX
PF      25-MAY-1995; 95US-0452592.
XX
PR      25-MAY-1993; 93US-0068948.
XX
PA      (USSA ) US SEC OF ARMY.
XX
PI      Kaplan DL, McGrath KP;
XX
DR      WPI; 1998-120029/11.
XX
PT      Production of peptide hetero:dimer(s) - in the fabrication of
        self-assembling protein-based structural material(s)

```

PS Disclosure; Fig 4C; 17pp; English.

XX
XX
CC AAM43024-28 represent recognition sequences obtained from the library of
CC recognition sequences coded for by AAV04943. This DNA sequence is
CC obtained by reverse translation. The codons were chosen to maximise
CC expression in Escherichia coli, and to introduce useful restriction for
CC subsequent genetic manipulations. The DNA sequence uses a "mixed site"
CC approach at the first base of the codons for amino acid positions 5 and
CC 7 of the heptad (see below). The formation of a self-assembling
CC structural material is mediated by the artificial recognition sequences
CC (e.g. present sequence). A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of
CC 2 different artificial peptide sequences, designed so as to
CC spontaneously heterodimerise. Both artificial peptides have a coiled-coil
CC alpha helical secondary structure having at least 2 heptad units. The
CC artificial peptides spontaneously heterodimerise into a self-assembled
CC protein-based structural material. The method is used to produce
CC nanoscale structural materials designed for specific functions e.g.
CC membranes, fibres, absorbents, reactive materials, etc. A specific
CC application is in the fabrication of selectively permeable membranes for
CC protective gloves and coated fabrics for fuel handlers and other
CC hazardous materials. Materials for medical implants, wound healants and
CC other medical treatments would also be feasible using this technology.

XA Sequence 84 AA:

XQ

```

Query Match          56.5%; Score 213; DB 19; Length 84;
Best Local Similarity 93.6%; Pred. No. 1.2e-15;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      16   GDKLNKVAQLRRKVRSLKDKAELKQEVSRLENETEDLKAKIGDINN 62
           |||||||:::|||||
Db       2   gdlknkvaglerevrsldkkaaelqevsrleenedlkakigdlkn 48

RESULT      5
AAMW43027
ID    AAMW43027 standard; Peptide: 84 AA.
XX
AC    AAMW43027;
XX
DT     08-MAY-1998 (first entry)
XX
DE    Artificial recognition sequence 4.
XX
KW    Recognition sequence; self-assembling; protein-based structural material;
KW    spontaneous heterodimerisation; coiled-coil alpha helix;
KW    selectively permeable membrane; coated fabric.
XX
OS    Synthetic.
XX
PN    US5712366-A.
XX
PD     27-JAN-1998.
XX
PF     25-MAY-1995; 95US-0452592.
XX
PR     25-MAY-1993; 93US-0068948.
XX
PA    (USSA ) US SEC OF ARMY.
XX
PI    Kaplan DL, McGrath KP;
XX
DR     WPI; 1998-120029/11.
XX
PT     Production of peptide hetero:dimer(s) - in the fabrication of
PT     self-assembling protein-based structural material(s)
XX
PS     Disclosure; Fig 4D; 17pp; English.
XX
CC    AAMW43024-28 represent recognition sequences obtained from the library of

```


7 of the heptad (see below). The formation of a self-assembling structural material is mediated by the artificial recognition sequences (e.g. present sequence). A method of producing a self-assembling protein-based structural material comprises mixing a multiplicity of 2 different artificial peptide sequences, designed so as to spontaneously heterodimerise. Both artificial peptides have a coiled-coil alpha helical secondary structure having at least 2 heptad units. The artificial peptides spontaneously heterodimerise into a self-assembled protein-based structural material. The method is used to produce nanoscale structural materials designed for specific functions e.g. membranes, fibres, absorbants, reactive materials, etc. A specific application is in the fabrication of selectively permeable membranes for protective gloves and coated fabrics for fuel handlers and other hazardous materials. Materials for medical implants, wound healants and other medical treatments would also be feasible using this technology.

Sequence 84 AA:

Query Match 52.3%; Score 197; DB 19; Length 84;
Best Local Similarity 85.1%; Pred. No. 6e-14; Indels 0; Gaps 0;
Matches 40; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKNKVAQLRRKRVSLDKRAELKQEVSRLENEIEDLAKKIGDLNN 62
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 2 gdlenevaqlerevrsledeaaelkqevsrleeneiedlkakigdlenn 48

RESULT 8
AAW43024
ID AAW43024 standard; Peptide; 84 AA.
XX
AC AAW43024;
XX
DT 08-MAY-1998 (first entry)
XX
DE Artificial recognition sequence 1.
XX
KM Recognition sequence: self-assembling; protein-based structural material;
KW spontaneous heterodimerisation; coiled-coil alpha helix;
KM selectively permeable membrane; coated fabric.
XX
OS Synthetic.
XX
PN US5712366-A.
XX
PD 27-JAN-1998.
XX
PF 25-MAY-1995; 95US-0452592.
XX
PR 25-MAY-1993; 93US-0068948.
XX
PA (USSA) US SEC OF ARMY.
XX
PI Kaplan DL, McGrath KP;
XX
DR WPI; 1998-120029/11.
XX
PT Production of peptide hetero:dimer(s) - in the fabrication of
XX self-assembling protein-based structural material(s)
XX
PS Disclosure; Fig 4A; 17pp; English.
XX
CC AAW43024-28 represent recognition sequences obtained from the library of
CC recognition sequences coded for by AAV04943. This DNA sequence is
CC obtained by reverse translation. The codons were chosen to maximise
CC expression in *Escherichia coli*, and to introduce useful restriction for
CC subsequent genetic manipulations. The DNA sequence use a "mixed site"
CC approach at the first base of the codons for amino acids positions 5 and
CC 7 of the heptad (see below). The formation of a self-assembling
CC structural material is mediated by the artificial recognition sequences
CC (e.g. present sequence). A method of producing a self-assembling
CC protein-based structural material comprises mixing a multiplicity of

2 different artificial peptide sequences, designed so as to spontaneously heterodimerise. Both artificial peptides have a coiled-coil alpha helical secondary structure having at least 2 heptad units. The artificial peptides spontaneously heterodimerise into a self-assembled protein-based structural material. The method is used to produce nanoscale structural materials designed for specific functions e.g. membranes, fibres, absorbants, reactive materials, etc. A specific application is in the fabrication of selectively permeable membranes for protective gloves and coated fabrics for fuel handlers and other hazardous materials. Materials for medical implants, wound healants and other medical treatments would also be feasible using this technology.

Sequence 84 AA:

Query Match 48.0%; Score 181; DB 19; Length 84;
Best Local Similarity 76.6%; Pred. No. 3e-12; Indels 0; Gaps 0;
Matches 36; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKNKVAQLRRKRVSLDKRAELKQEVSRLENEIEDLAKKIGDLNN 62
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 2 gdlenevaqlerevrsledeaaelkqevsrleeneiedlkakigdlenn 48

RESULT 9
AAB10557
ID AAB10557 standard; protein; 42 AA.
XX
AC AAB10557;
XX
DT 15-DEC-2000 (first entry)
XX
DE Coiled coil protein A1.
XX
KM Coiled coil protein; copolymer blocker; vulnerrary; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KM antibiotic; dressing; abrasion; burn; A1.
XX
OS Unidentified.
XX
PN US6090911-A.
XX
PD 18-JUL-2000.
XX
PF 22-OCT-1997; 97US-0956307.
XX
PR 22-OCT-1997; 97US-0956307.
XX
PA (UYMA-) UNIV MASSACHUSETTS.
XX
PI Petka WA, McGrath KP, Tirrell DA;
XX
DR WPI; 2000-542235/49.
XX
PT Synthetic block copolymer with an antibiotic compound, useful for
XX dressing abrasion, burn or non-puncture wound, comprises two
XX alpha-helical protein blocks, random-coil protein block and linker
XX proteins -
XX
PS Example 1; Column 25-26; 53pp; English.
XX
CC This invention describes a novel synthetic block copolymer XYZ comprising
CC two alpha-helical protein blocks X and Z, a water soluble, random-coil
CC protein block Y which links X and Z, and linker proteins for linking the
CC alpha-helical protein blocks to the random-coil protein block. The
CC products of the invention have vulnerrary activity. A copolymer produced
CC from a gene that encodes a derivative of AC_1_0A copolymer was suspended
CC in water to form a gel, which was then used to treat a wound. The cell
CC binding domain was the integrin ArgGlyAspser sequence which is known to
CC bind to alpha5beta1 proteins expressed on fibroblast cells necessary for
CC matrix formation for the regeneration of skin. Fibroblasts become
CC entrapped within the gel and thus remain at the site of the wound,
CC serving as a scaffold for the regeneration of tissue surrounding the

CC wound. A synthetic block copolymer together with an antibiotic compound
CC is useful for wound dressing. The copolymer is useful for dressing
CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the
CC polymers of equal size are produced biologically from a single template.
CC The intermediate binding of alpha-helical blocks are monodisperse which
CC gives a uniform pore size depending on the length of random-coil block.
CC The new copolymers have low molecular weight hence have decreased
CC viscosity solutions or gels under suitable conditions. This sequence
CC represents the coiled coil protein A1 which is used to illustrate the
CC method of the invention.

XX Sequence 42 AA:

Query Match 42.2%; Score 159; DB 21; Length 42;

Best Local Similarity 76.2%; Pred. No. 2, 9e-10;
Matches 33; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY 15 GDLKKNVAQLKRRVRSLSKDKAELEKQVSRLENEDLKAK 56
Db 1 sgdlenevaqlerevrsrldeaealeqkvrsrlkneidlkae 42

RESULT 10

AAW43029 ID AAW43029 standard; Peptide: 42 AA.

XX AC AAW43029;

XX DT 08-MAY-1998 (first entry)

XX DE Amino acid sequence giving a library of recognition sequences.

XX KM Recognition sequence: self-assembling; protein-based structural material;
XX spontaneous heterodimerization; coiled-coil alpha helix;
XX selectively permeable membrane; coated fabric.

XX OS Synthetic.

XX FT Key Location/Qualifiers

FT MISC-difference 5 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 7 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 12 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 14 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 19 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 21 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 26 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 28 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 33 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 35 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 40 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 42 /note= "any charged residue, preferably Glu or Lys"
FT MISC-difference 42 /note= "any charged residue, preferably Glu or Lys"

XX US5712366-A.

XX PD 27-JAN-1998.

XX PE 25-MAY-1995; 95US-0452592.

XX PR 25-MAY-1993; 93US-0068948.

PA (USSA) US SEC OF ARMY.

XX PI Kaplan DL, McGrath KP;

XX DR WPI: 1998-120029/11.

XX DR N-PSDB; AAW04943.

PT Production of peptide hetero: dimer(s) - in the fabrication of
PT self-assembling protein-based structural material(s)

XX PS Claim 2; Columns 13-14; 17pp; English.

XX CC The present sequence is used to obtain an entire library of recognition
XX sequences. The DNA sequence encoding the present sequence is obtained by
XX reverse translation. The codons were chosen to maximize expression in
XX Escherichia coli, and to introduce useful restriction for subsequent
XX genetic manipulations. The DNA sequence use a "mixed site" approach at
XX the first base of the codons for amino acids positions 5 and 7 of the
XX heptad (see below). The formation of a self-assembling structural
XX material is mediated by the artificial recognition sequences obtained
XX from the present sequence. A method of producing a self-assembling
XX protein-based structural material comprises mixing a multiplicity of
XX 2 different artificial peptide sequences, designed so as to spontaneously
XX heterodimerise. Both artificial peptides have a coiled-coil alpha helical
XX secondary structure having at least 2 heptad units. The artificial
XX peptides spontaneously heterodimerise into a self-assembled protein-based
XX structural material. The method is used to produce nanoscale structural
XX materials designed for specific functions e.g. "membranes, fibres",
XX absorbants, reactive materials, etc. A specific application is in the
XX fabrication of selectively permeable membranes for protective gloves and
XX coated fabrics for fuel handlers and other hazardous materials. Materials
XX for medical implants, wound healants and other medical treatments would
XX also be feasible using this technology.

XX SQ Sequence 42 AA:

Query Match 32.9%; Score 124; DB 19; Length 42;

Best Local Similarity 72.5%; Pred. No. 1, 4e-06;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 16 GDLKKNVAQLKRRVRSLSKDKAELEKQVSRLENEDLKAK 55
Db 2 gdlxnvaqlrxrvrsrlxkaaelxqvxrsrlxnxdlxa 41

RESULT 11

AAB08364 ID AAB08364 standard; Peptide: 43 AA.

XX AC AAB08364;

XX DT 20-DEC-2000 (first entry)

XX DE Amino acid sequence of a coiled-coil peptide.

XX KW Binding partner; protein modification; post-translational modification;
XX modulator; coiled-coil structure.

XX OS Unidentified.

XX FT Key Location/Qualifiers

FT CLEAVAGE-site 24 /note= "thrombin cleavage site"
FT MISC-difference 39 /note= "fluorophore attachment site"

XX WO200050902-A2.

XX PD 31-AUG-2000.

XX PE 25-FEB-2000; 2000WO-GB00669.

PR 25-FEB-1999; 99GB-0004398.
PA (FLUO-) FLUORESCIENCE LTD.
PI Colyer J, Craig RK, Maschio A, Mezna M;
XX WPI; 2000-572119/53.
DR
XX
XX
PT High throughput assay for monitoring modification of polypeptides and
PT modification of the modifications
PS Disclosure; Page 61; 128pp; English.
XX
CC The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC of state of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The
CC polypeptides, support and methods can be used to analyse a sample to
CC determine if modification of a polypeptide is taking place and to
CC identify modulators of the modification. This is useful for monitoring
CC the post-translational modification of proteins. AAB08364-65 represent
CC binding partners which have coiled-coil structures, and may be used in
CC the method of the invention to assay for thrombin activity.
CC
XX
SQ Sequence 43 AA;

Query Match 32.1%; Score 121; DB 21; Length 43;
Best Local Similarity 46.7%; Pred. No. 3.1e-06;
Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

OY 5 HHHHHGMSGDLKNVQOLKRVRSIKDKAELKQVSRLENE 49
Db 1 hhhhhngsla-qlegeladqegenrqldeqeladqlegetakleage 43

RESULT 12
AAB08383
ID AAB08383 standard; peptide; 43 AA.
XX
AC AAB08383;
XX
DT 20-DEC-2000 (first entry)
XX
DE Peptide used to assay for tobacco etch virus protease activity.
XX
XX Binding partner; protein modification; post-translational modification;
KW modulator; coiled-coil structure.
XX
OS Unidentified.
XX
PN WO200050902-A2.
XX
PD 31-AUG-2000.
XX
PF 25-FEB-2000; 2000MO-GB00669.
XX
PR 25-FEB-1999; 99GB-0004398.
PA (FLUO-) FLUORESCIENCE LTD.
PI Colyer J, Craig RK, Maschio A, Mezna M;
XX WPI; 2000-572119/53.
DR
XX High throughput assay for monitoring modification of polypeptides and
PT modification of the modifications
XX
PS Example 9; Page 80; 128pp; English.

XX
CC The specification describes a method for analysing a sample. The method
CC comprises immobilising a polypeptide to a physical support, contacting
CC the immobilised polypeptide with a test sample which may contain an
CC agent capable of modifying the immobilised polypeptide, contacting the
CC immobilised polypeptide with a binding partner polypeptide, where
CC association of both polypeptide is dependent on the modification
CC of state of the immobilised polypeptide, and measuring the association of
CC the binding partner polypeptide to the immobilised polypeptide. The
CC polypeptides, support and methods can be used to analyse a sample to
CC determine if modification of a polypeptide is taking place and to
CC identify modulators of the modification. This is useful for monitoring
CC the post-translational modification of proteins. AAB08383-84 represent
CC binding partners which are used, in the method of the invention, to
CC assay for tobacco etch virus (TEV) protease activity.
CC
XX
SQ Sequence 43 AA;

Query Match 32.1%; Score 121; DB 21; Length 43;
Best Local Similarity 46.7%; Pred. No. 3.1e-06;
Matches 21; Conservative 14; Mismatches 8; Indels 2; Gaps 1;

OY 5 HHHHHGMSGDLKNVQOLKRVRSIKDKAELKQVSRLENE 49
Db 1 hhhhhngsla-qlegeladqegenrqldeqeladqlegetakleage 43

RESULT 13
AAB70797
ID AAB70797 standard; Protein; 103 AA.
XX
AC AAB70797;
XX
DT 25-MAY-2001 (first entry)
XX
DE HIV protease-associated protein #2.
XX
XX Protease substrate; detection; cleavage site; stretched helix;
KW screening.
XX
OS Human immunodeficiency virus.
XX
PN DE19941239-A1.
XX
PD 01-MAR-2001.
XX
PF 26-AUG-1999; 99DE-1041239.
XX
PR 26-AUG-1999; 99DE-1041239.
PA (IMB-) IMB INST MOLEKULARE BIOTECHNOLOGIE EV.
PI Aldinger U, Hill O, Hallsch A, Steirnuecke P;
XX WPI; 2001-246019/26.
DR
XX
PT Protease substrate useful for measuring protease activities and in
PT online screening assays for proteases or protease inhibitors, comprises
PT a polypeptide with a binding site, a cleavage site and a stretched
PT helix -
XX
PS Disclosure; Fig 2; 14pp; German.
XX
CC This invention describes a novel protease substrate for detecting a
CC target protease which comprises a polypeptide containing a high-affinity
CC binding site (1) for binding to a support, a cleavage site (2) specific
CC for the target protease, and a sequence of amino acids forming a
CC stretched helix (3). The substrate is useful for measuring protease
CC activities and in online screening assays for proteases or protease
CC inhibitors, especially based on surface plasmon resonance.
XX
SQ Sequence 103 AA;

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:46 ; Search time 48.42 Seconds
(without alignments)
30.787 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377
Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRRAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 segs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *
5: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
6: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	279.5	74.1	76	3	US-08-956-307B-18
2	243.5	64.6	76	3	US-08-956-307B-17
3	221	58.6	84	1	US-08-452-592B-7
4	213	56.5	84	1	US-08-452-592B-8
5	205	54.4	84	1	US-08-452-592B-9
6	199	52.8	42	3	US-08-956-307B-2
7	197	52.3	84	1	US-08-452-592B-10
8	181	48.0	84	1	US-08-452-592B-6
9	159	42.2	42	3	US-08-956-307B-1
10	124	32.9	42	3	US-08-452-592B-11
11	91	24.1	18	3	US-08-956-307B-4
12	87	23.1	14	3	US-08-956-307B-3
13	82.5	21.9	110	1	US-08-434-705B-15
14	82.5	21.9	110	2	US-08-086-201-15
15	82	21.8	18	1	US-08-578-649-14
16	81	21.5	42	5	PCT-US94-06655-9
17	81	21.5	48	1	US-08-457-245-21
18	81	21.5	126	1	US-08-268-348A-12
19	81	21.5	133	1	US-08-268-348A-8
20	81	21.5	133	1	US-08-268-348A-10
21	79.5	21.1	31	2	US-08-662-227-37
22	79	21.0	32	3	US-08-737-336-1
23	79	21.0	39	2	US-08-679-865-39
24	79	21.0	39	2	US-08-680-876-39
25	79	21.0	69	2	US-08-687-865A-21
26	79	21.0	69	4	US-09-043-711-21
27	79	21.0	88	2	US-08-690-011A-11

28	79	21.0	254	4	US-09-004-731-89	Sequence 89, Appl
29	79	21.0	254	4	US-09-004-731-92	Sequence 92, Appl
30	79	21.0	254	4	US-08-749-699-89	Sequence 89, Appl
31	79	21.0	254	4	US-08-749-699-92	Sequence 92, Appl
32	79	21.0	351	2	US-08-701-191A-3	Sequence 3, Appl1
33	79	21.0	532	3	US-08-737-336-6	Sequence 6, Appl1
34	79	21.0	631	1	US-08-605-541B-12	Sequence 12, Appl
35	78.5	20.8	98	2	US-08-690-011A-19	Sequence 19, Appl
36	78	20.7	16	1	US-08-578-649-13	Sequence 13, Appl
37	78	20.7	20	4	US-08-640-737-32	Sequence 32, Appl
38	78	20.7	35	2	US-08-491-527A-2	Sequence 2, Appl1
39	78	20.7	35	2	US-08-845-659-2	Sequence 2, Appl1
40	78	20.7	35	3	US-08-845-658-2	Sequence 2, Appl1
41	78	20.7	122	3	US-08-956-307B-19	Sequence 19, Appl
42	78	20.7	184	3	US-09-023-082A-30	Sequence 30, Appl
43	75.5	20.0	153	4	US-08-676-318A-34	Sequence 34, Appl
44	75	19.9	12	5	PCT-US94-09700-4	Sequence 4, Appl1
45	75	19.9	20	2	US-08-543-020-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-956-307B-18
Sequence 18, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Tirrell, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ. ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-956-307B-18

Query Match 74.1%; Score 279.5; DB 3; Length 76;
Best Local Similarity 89.1%; Pred. No. 1,7e-23;
Matches 57; Conservative 1; Mismatches 1; Indels 5; Gaps 1;
OY 1 MRGSHHHHGS-----MASGDLKNKYAQLKRYKRSLKDKRAAEKQVSRLENEIEDLKA 55

Db 1 MGSHHHHHSGDDDKKMSGDLKKNVQQLKRRVSLKDKAAELKQEVSRLENEIDLKA 60
QY 56 KIGD 59
61 KIGD 64

RESULT 2

US-08-956-307B-17
; Sequence 17, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-8906
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-956-307B-17

Query Match 64.6%; Score 243.5; DB 3; Length 76;
Best Local Similarity 75.0%; Pred. No. 1.3e-19;

Matches 48; Conservative 10; Mismatches 1; Indels 5; Gaps 1;

QY 1 MGSHHHHHSGS-----MASGDLKKNVQQLKRRVSLKDKAAELKQEVSRLENEIDLKA 55
Db 1 MGSHHHHHSGDDDKKMSGDLKKNVQQLKRRVSLKDKAAELKQEVSRLENEIDLKA 60

QY 56 KIGD 59
61 KIGD 64

RESULT 3

US-08-452-592B-7
; Sequence 7, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.

; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS 6.22
SOFTWARE: Wordperfect Version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452.592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068.948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Ranucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Ranucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510
TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-452-592B-7

Query Match 58.6%; Score 221; DB 1; Length 84;
Best Local Similarity 97.9%; Pred. No. 3.7e-17;

Matches 46; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 GDLKKNVQQLKRRVSLKDKAAELKQEVSRLENEIDLKA KIGD LNN 62
Db 2 GDLKKNVQQLKRRVSLKDKAAELKQEVSRLENEIDLKA KIGD LKN 48

RESULT 4

US-08-452-592B-8
; Sequence 8, Application US/08452592B
; Patent No. 5712366

; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.

; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using

; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSC-CC (Patent Counsel)

; CITY: Natick
; STATE: Massachusetts

; COUNTRY: United States of America
; ZIP: 01760-5035

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS 6.22
SOFTWARE: Wordperfect Version 6.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Ramucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Ramucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510
TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-452-592B-8

Query Match 56.5%; Score 213; DB 1; Length 84;
Best Local Similarity 93.6%; Pred. No. 2.7e-16;
Matches 44; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKNVVAOLKRRVSLKDKAAELKOEVSRLNEIEDLAKIGDLNN 62
DB 2 GDLKNVVAOLEREVRSLLKDKAAELKOEVSRLNEIEDLAKIGDLNN 48

RESULT 5
US-08-452-592B-9
Sequence 9, Application US/08452592B
Patent No. 5712366
GENERAL INFORMATION:
APPLICANT: McGrath, Kevin P.
APPLICANT: Kaplan, David L.
TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
TITLE OF INVENTION: Self-Assembling Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: U.S. Army Soldier Systems Command
STREET: AMSCC-CC (Patent Counsel)
CITY: Natick
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01760-5035
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS 6.22
SOFTWARE: WordPerfect Version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Ramucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Ramucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510

TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-452-592B-9

Query Match 54.4%; Score 205; DB 1; Length 84;
Best Local Similarity 89.4%; Pred. No. 1.9e-15;
Matches 42; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 16 GDLKNVVAOLKRRVSLKDKAAELKOEVSRLNEIEDLAKIGDLNN 62
DB 2 GDLNEVVAOLEREVRSLLKDKAAELKOEVSRLNEIEDLAKIGDLNN 48

RESULT 6
US-08-956-307B-2
Sequence 2, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Titrill, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-956-307B-2

Query Match 52.8%; Score 199; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 3.8e-15;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 15 SGDLKNVVAOLKRRVSLKDKAAELKOEVSRLNEIEDLAK 56
DB 1 SGDLKNVVAOLKRRVSLKDKAAELKOEVSRLNEIEDLAK 42

RESULT 7

```

US-08-452-592B-10
: Sequence 10, Application US/08452592B
: Patent No. 5712366
:
: GENERAL INFORMATION:
:   APPLICANT: McGrath, Kevin P.
:   APPLICANT: Kaplan, David L.
:   TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
:   TITLE OF INVENTION: Self-Assembling Proteins
:   NUMBER OF SEQUENCES: 11
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: U.S. Army Soldier Systems Command
:   STREET: AMSCC-CC (Patent Counsel)
:   CITY: Natick
:   STATE: Massachusetts
:   COUNTRY: United States of America
:   ZIP: 01760-5035
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
:   COMPUTER: IBM PC Compatible
:   OPERATING SYSTEM: DOS 6.22
:   SOFTWARE: Wordperfect Version 6.1
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/452,592B
:   FILING DATE: 25 May 1995
:   CLASSIFICATION: 530
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 08/068,948
:   FILING DATE: 25 May 1993
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Donahue, Richard J.
:   NAME: Lamming, John H.
:   NAME: Rannucci, Vincent J.
:   REGISTRATION NUMBER: 22,062 (Donahue)
:   REGISTRATION NUMBER: 34,857 (Lamming)
:   REGISTRATION NUMBER: 29,579 (Rannucci)
:   REFERENCE/DOCKET NUMBER: NA-1096D
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 508-233-4510
:   TELEFAX: 508-233-5167
:   INFORMATION FOR SEQ ID NO: 10:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 84 amino acids
:     TYPE: amino acid
:     TOPOLOGY: linear
:
: US-08-452-592B-10
:
: Query Match          52.3%; Score 197; DB 1; Length 84;
: Best Local Similarity 85.1%; Pred. No. 1,4e-14;
: Matches 40; Conservative 6; Mismatches 1; Indels 0
:
: Oy      16  GDLKKNVLAOLKRRVRSLEDKRAELKQVSRLENFIEDIKAKIGDLNN 62
:           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
: Db      2  GDLENVLAQLEREVSRLSEDAAEIKQVSRLENFIEDIKAKIGDLN 48
:
: RESULT      8
: US-08-452-592B-6
:   Sequence 6, Application US/08452592B
:   Patent No. 5712366
:   GENERAL INFORMATION:
:     APPLICANT: McGrath, Kevin P.
:     APPLICANT: Kaplan, David L.
:     TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
:     TITLE OF INVENTION: Self-Assembling Proteins
:     NUMBER OF SEQUENCES: 11
:     CORRESPONDENCE ADDRESS:
:     ADDRESSEE: U.S. Army Soldier Systems Command
:     STREET: AMSCC-CC (Patent Counsel)
:     CITY: Natick
:     STATE: Massachusetts
:     COUNTRY: United States of America
:     ZIP: 01760-5035

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: DOS 6.22
SOFTWARE: Wordperfect Version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,592B
FILING DATE: 25 May 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/068,948
FILING DATE: 25 May 1993
ATTORNEY/AGENT INFORMATION:
NAME: Donahue, Richard J.
NAME: Lamming, John H.
NAME: Rannucci, Vincent J.
REGISTRATION NUMBER: 22,062 (Donahue)
REGISTRATION NUMBER: 34,857 (Lamming)
REGISTRATION NUMBER: 29,579 (Rannucci)
REFERENCE/DOCKET NUMBER: NA-1096D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 508-233-4510
TELEFAX: 508-233-5167
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 84 amino acids
TYPE: amino acid
TOPOLOGY: linear

Query Match	48.08;	Score 181;	DB 1;	Length 84;
Best Local Similarity	76.68;	Pred. No. 7.3e-13;		
Matches 36;	Conservative 10;	Mismatches 1;	Indels 0;	Gaps 0;

```
QY      16 GDLKNNVAQLKRRVRSLLDKAAELQEVSRLENEIEDLKAKIGD LNN 62
      |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      2  GDLLENVAQLEREVRSLEDEAAELQKVSRLKNEIEDLKAEIGD LLEN 48
```

US RESULT 9
US-08-956-307B-1
Sequence 1, Application US/08956307B
Patent No. 6090911
GENERAL INFORMATION:
APPLICANT: Petka, Wendy A.
APPLICANT: Tirrell, David A.
APPLICANT: Kevin P. McGrath
TITLE OF INVENTION: REVERSIBLE HYDROGELS
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,307B
FILING DATE: 22-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07880/033001
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-1

Query Match          42.2%; Score 159; DB 3; Length 42;
Best Local Similarity 76.2%; Pred. No. 7.4e-11;
Matches 32; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 15 SGD LKNKVAQLKRRVSLKDKAELKQEVSRLENEIEDLKAK 56
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 SGDLENEVQAQLEREVRSLDEDAELKQVSRLENEIEDLKAE 42

RESULT 10
US-08-452-592B-11
; Sequence 11, Application US/08452592B
; Patent No. 5712366
; GENERAL INFORMATION:
; APPLICANT: McGrath, Kevin P.
; APPLICANT: Kaplan, David L.
; TITLE OF INVENTION: Fabrication of Nanoscale Materials Using
; TITLE OF INVENTION: Self-Assembling Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. Army Soldier Systems Command
; STREET: AMSCC-CC (Patent Counsel)
; CITY: Natick
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01760-5035
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: DOS 6.22
; SOFTWARE: Wordperfect Version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,592B
; FILING DATE: 25 May 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/068,948
; FILING DATE: 25 May 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Donahue, Richard J.
; NAME: Lamming, John H.
; NAME: Ranucci, Vincent J.
; REGISTRATION NUMBER: 22,062 (Donahue)
; REGISTRATION NUMBER: 34,857 (Lamming)
; REGISTRATION NUMBER: 29,579 (Ranucci)
; REFERENCE/DOCKET NUMBER: NA-1096D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-233-4510
; TELEFAX: 508-233-5167
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-452-592B-11

Query Match          32.9%; Score 124; DB 1; Length 42;
Best Local Similarity 72.5%; Pred. No. 4.2e-07;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```

QY 16 GDLKKNVAQLKRRVSLKDKAELKQEVSRLENEIEDLKAK 55
    ||| | |||| | |||| | |||| | |||| | |||| |
Db 2 GDLKKNVAQLKRRVSLKDKAELKQVSRLENEIEDLKAK 41

RESULT 11
US-08-956-307B-4
; Sequence 4, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-956-307B-4

Query Match          24.1%; Score 91; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 0.00054;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ISGLNNTSGIRPPAKLN 74
    |||||||:|||||:|||||:|||||:|||||:
Db 1 ISGLNNTSGIRPPAKLN 18

RESULT 12
US-08-956-307B-3
; Sequence 3, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
```



```

QY      1  MRSNNNNNN--GSMASGDYKKNVAQKRYK---VRSJDKKAELQFVSSTLENNEDLKA 55
        |  |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      1  MGSNNNNNNSSSLVPRGSMQEDAEALPRRALDYTSANVDQASNEKELIELALDVELKLKS 600
        |
QY      56  K  56
        |
Db      61  K  61

```

RESULT 15
US-08-578-649-14

: Sequence 14, Application US/08578649
: Patent No. 5770366

Patent No. 3770366
; GENERAL INFORMATION
:

; GENERAL INFORMATION:

APPLICANT: Ulrich Bogdan

APPLICANT: Relnhard Butner
APPLICANT: Br1414to Kaluza

APPLICANT: BRIGITTE KALUZA
TITLE OF INVENTION: MELANOM

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Felfe & L
STREET: 905 Third Ave

STREET: 805 Third Avenue
CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE
MEDIUM TYPE.

COMPUTER: IBM

```

;
; COMPILE: ADM.FS/2
; OPERATING SYSTEM: PC

```

```
SOFTWARE: wordperfect
```

CURRENT APPLICATION D

APPLICATION NUMBER: US/0
FILING DATE: 29-Jul-1999

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE P 43 24 24

FILING DATE: 20-July-1993

NAME: Andrew L. Tlato

NAME: ANDREW L. LIAJ
REGISTRATION NUMBER:

REFERENCE/DOCKET NUMBER: BOER 1035-PFF/ALTT

TELECOMMUNICATION INFORMATION ;
ELECTRONIC 0179 608 0000

TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884

TELEFAX: (212) 838-3884
INFORMATION FOR SEO ID NO: 14:

1 INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

LENGTH: 18 amino acids

TYPE: amino acid

```

; STRANDEDNESS: single
; TOPOLOGY: linear

```

TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-578-649-14

Query Match 21

Query match	21
Best local similarity	76

Matches 13; Conservatively

1000

Query Match	21.8%	Score 82;	DB 1;	Length 18;
Best Local Similarity	76.5%	Pred. No. 0.005;		
Matches 13;	Conservative 1;	Mismatches 3;	Indels 0;	Gaps 0;
OY	1	MRGSHHHHRCGSMASGD	17	
Db	1	MRGSHHHHRCGSVDDDD	17	

Search completed: July 3, 2001, 14:56:47
Job time: 537 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:59:46 ; Search time 63.58 Seconds
(without alignments)
88,659 Million cell updates/sec

Title: US-09-490-291-11

Perfect score: 377
Sequence: 1 MRGSHHHHHGSMASGDLKN.....AKIGDLNNTSGIRPAPKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	25.1	1186	2	G69708 chromosome segrega
2	89	23.6	315	3	JC7572 somite Maf1 protei
3	88	23.3	259	2	A70359 hydrogenase expres
4	86	22.8	155	2	I50167 B-G antigen - chic
5	85.5	22.7	323	2	I49529 transcription fact
6	85	22.5	321	2	D39371 Ig V-region-like B
7	85	22.5	880	2	D39371 conserved hypochet
8	85	22.5	979	2	J00894 P115 protein - MYC
9	84.5	22.4	281	2	F75216 hypothetical prote
10	84	22.3	1199	2	T29145 hypothetical prote
11	82	21.8	173	2	JC5610 tropoin I - sea s
12	82	21.8	311	2	A56235 transcription acti
13	82	21.8	766	2	S37894 hypothetical prote
14	82	21.8	1188	2	S37894 chromosome segrega
15	81.5	21.6	212	2	H69402 hypothetical prote
16	81	21.5	166	2	S73342 hypothetical prote
17	81	21.5	270	2	G71061 hypothetical prote
18	81	21.5	396	2	T41405 hypothetical prote
19	81	21.5	409	2	F72504 probable 26S prote
20	80.5	21.4	269	2	T18335 lcmg protein - Leg
21	80.5	21.4	647	2	A84265 hypothetical prote
22	80	21.2	102	2	F64341 hypothetical prote
23	80	21.2	103	2	B60608 myosin heavy chain
24	80	21.2	527	2	S33068 myosin heavy chain
25	80	21.2	1225	2	A49464 chromosome segrega
26	80	21.2	1940	2	A59287 myosin heavy chain
27	79.5	21.1	565	1	HMIVE3 hemagglutinin prec
28	79.5	21.1	565	1	HMIVE4 hemagglutinin prec
29	79	21.0	318	2	T49167 hypothetical prote

30	79	21.0	369	1	TVFVAF transforming prote
31	79	21.0	407	2	F72343 hypothetical prote
32	79	21.0	554	2	G72361 hypothetical prote
33	78.5	20.8	564	1	HMIVE2 hemagglutinin prec
34	78.5	20.8	565	1	HMIVE2 hemagglutinin prec
35	78.5	20.8	1956	2	T16416 hypothetical prote
36	78	20.7	456	2	E86903 hypothetical prote
37	78	20.7	1169	2	A64505 P115 homolog - Met
38	78	20.7	1313	2	A48467 myosin heavy chain
39	78	20.7	1938	1	A40997 myosin heavy chain
40	77.5	20.6	348	2	T04618 heat shock protein
41	77.5	20.6	778	2	T30430 hypothetical prote
42	77.5	20.6	1556	2	F96587 hypothetical prote
43	77	20.4	292	2	I51171 transcription fact
44	77	20.4	338	2	I38587 retrovirus-related
45	77	20.4	1046	2	T42720 cytoplasmic linker

ALIGNMENTS

RESULT 1

G69708 Chromosome segregation SMC protein - Bacillus subtilis

N/Alternate names: multichromosome stabilizing protein SMC

C/Species: Bacillus subtilis

C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C/Accession: G69708; J04819; PC4029

R/Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Allion, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerson, P.T.; Enlian, K.D.; Errington, J.; Farell, C.; Ferrari, A.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galluzzi, A.; Gal lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Lardino Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidos, A.; Lardino A.; Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portee Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A/Reference number: A69580; M01D:98044033

A/Accession: G69708

A/Molecule type: DNA

A/Residues: 11186 <KUN>

A/Cross-references: GB:299112; GB:AL009126; NID:92633902; PIDN:CA13467.1; PID:926339

A/Experimental source: strain 168

R/Oguro, A.; Kakeshita, H.; Takamatsu, H.; Nakamura, K.; Yamane, K.

Gene 172, 17-24, 1996

A/Title: The effect of Srb, a homologue of the mammalian SRP receptor alpha-subunit,

A/Reference number: J04819; M01D:96257247

A/Accession: J04819

A/Molecule type: DNA

A/Residues: 1-49, 'G', '51-161, 'G', '163-174, 'E', '176-177, 'G', '179-191, 'G', '193-227, 'P', '229-2 74, 'P', '476-493, 'D', '495-514, 'D', '516-541, 'V', '543-545, 'P', '547-585, 'SKPLRGSGPAFTISF', '601 'E', '741-1186 <OGU>

A/Cross-references: DDBJ:D04116; NID:91389548; PIDN:BA10977.1; PID:91237015

R/Oguro, A.; Kakeshita, H.; Honda, K.; Takamatsu, H.; Nakamura, K.; Yamane, K.

DNA Res. 2, 95-100, 1995

A/Title: srb: a Bacillus subtilis gene encoding a homologue of the alpha-subunit of t

A/Reference number: J04093; M01D:96093930

A/Accession: PC4029

A/Molecule type: DNA

A/Residues: 1171-1186 <OG2>

A/Cross-references: DDBJ:D49781; NID:91237017; PIDN:BA08615.1; PID:91237018

C/Genetics: A; Gene: smc

C/Superfamily: chromosome segregation protein SMC1

```
Query Match 25.1%, Score 94.5; DB 2; Length 1186;
Best Local Similarity 36.8%; Pred. No. 2.2;
Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;

OY 11 GSMASGDLKNKYAOLKRRVRSIKD-----KAELEKQVSRLENEIEDLKAKIGDL 60
DB 658 GSMGGAAVKKKNNLSILGSRLEEDVTKRLAEKMEKTKALLEDEVKTKHSDIMDKMLADL 717

OY 61 NNT-SGIR 67
DB 718 RETGEGLR 725

RESULT 2
JC7572
somite Maf1 protein - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C:Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C:Accession: JC7572
R:Kajihara, M.; Kawachi, S.; Kobayashi, M.; Ogino, H.; Takahashi, S.; Yasuda, K.
J. Biochem. 129, 139-146, 2001
A:Title: Isolation, characterization, and expression analysis of zebrafish large Mafs.
A:Reference number: JC7572; MUID: 21064923
A:Accession: JC7572
A:Molecule type: mRNA
A:Residues: 1-315 <KAU>
A:Cross-references: DDBJ:AB006324
C:Comment: This protein, a member of large Maf protein subfamily, containing a basic-leu
important roles in somitogenesis, and has specific functions as to the determination and
C:Genetics:
A:Gene: s-maf1

Query Match 23.6%; Score 89; DB 3; Length 315;
Best Local Similarity 25.5%; Pred. No. 1.7;
Matches 24; Conservative 12; Mismatches 20; Indels 38; Gaps 2;

OY 5 HHHHHGSMASGDL-----KNKVAOLKRRVRSIKDKAAE- 38
DB 179 HHHHHGHNHNAHRLDRSDEQLVMTVRELNRQLRGFSKEVEYTLKQKRRRLKKNRGYAQ 238

OY 39 -----LKQVSRLENEIEDLKAKIGDL 60
DB 239 SCRYKRVQORHMLESEKCTLOQVBDLQDVARL 272

RESULT 3
A70359
hydrogenase expression/formation protein B - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 24-Sep-1999
C:Accession: A70359
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'
V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID: 98196666
A:Accession: A70359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <AAQ>
A:Cross-references: GB:AE000701, NID:g2983260; PIDN:AA06680.1; PID:g2983266; GB:AE00065
A:Experimental source: strain VF5
C:Genetics:
A:Gene: hypB
C:Superfamily: hydrogenase expression/formation protein hypB

Query Match 23.3%; Score 88; DB 2; Length 259;
Best Local Similarity 28.6%; Pred. No. 1.7;
Matches 26; Conservative 12; Mismatches 29; Indels 24; Gaps 3;

OY 5 HHHHHGSMASGDLKNKYAOLKRRVRSIKDKAAELKQ-----EVS 44
```

```
DB 19 HHHHHHTPALGD--KRTVEYVKTKLSANDQAESNRHFERHGLAVALMSSPESGKTT 76
OY 45 RLENEIEDLK--AKIGDLNNTSGIRPAKL 73
DB 77 LLERTIELLKDELKIGVTEGDLFTNRDAERI 107

RESULT 4
I50167
B-G antigen - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I50167
R:Kaufman, J.; Salomonsen, J.; Skjodet, K.
Immunogenetics 30, 440-451, 1989
A:Title: B-G cDNA clones have multiple small repeats and hybridize to both chicken mH
A:Reference number: I50161; MUID:90077532
A:Accession: I50167
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-155 <KAU>
A:Cross-references: GB:M27668; NID:g211253; PIDN:AAA48624.1; PID:g211254

Query Match 22.8%; Score 86; DB 2; Length 155;
Best Local Similarity 27.5%; Pred. No. 1.5;
Matches 22; Conservative 17; Mismatches 23; Indels 18; Gaps 2;

OY 7 HHHHSMASGDLKNKYA-----QLKRVVSLDKAAELQVSRLENEIED 52
DB 19 HCNEMDLSAADLKQIALVEQREGVEEMHTQLRKHYEKLASRAANLTQLKLENEIEE 78

OY 53 LKAKIGDLNNTSGIRPAK 72
DB 79 VEKHL-----KITGRAPNIK 94

RESULT 5
I49529
transcription factor-kr - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 16-Jul-1999
C:Accession: I49529
R:Corde, S.P.; Barsh, G.S.
Cell 79, 1025-1034, 1994
A:Title: The mouse segmentation gene kr encodes a novel basic domain-leucine zipper t
A:Reference number: A55200; MUID:95094266
A:Accession: I49529
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-323 <RES>
A:Cross-references: GB:I36435; NID:g625043; PIDN:AAA65689.1; PID:g625044
C:Superfamily: maf transforming protein; maf homology
C:Keywords: leucine zipper; transcription factor
F:212-301/Domain: maf homology <MAF>

Query Match 22.7%; Score 85.5; DB 2; Length 323;
Best Local Similarity 23.1%; Pred. No. 3.4;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

OY 5 HHHHHN-----GSMAS----- 15
DB 162 HHHHHNQASPPSSAASPAQQLPTSHPGGRPATATAAGGNSVEDRFSDDQLVMSV 221
OY 16 -----GDLKNKYAOLKRRVRSIKDK-----AAELK 40
DB 222 RELNRLNLGFTGDEYVRLKQKRRRLKKNRGYASCRYYKQKHNLENEKTOILQOVEDLK 281
OY 41 QEVSRLENEIEDLKAKIGDLNNTSGIR 67
DB 282 DEVSRLEARERDAKYKCEKLAN-SGFR 307
```

```
RESULT 6
Ig V-region-like B-G antigen 17 - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 20-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 21-Jul-2000
C:Accession: D39371
R:Miller, M.M.; Goto, R.; Young, S.; Chirivella, J.; Hawke, D.; Miyada, C.G.
Proc. Natl. Acad. Sci. U.S.A. 88, 4377-4381, 1991
A:Title: Immunoglobulin variable-region-like domains of diverse sequence within the major
A:Reference number: A39371; MIMD:91239571
A:Accession: D39371
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-321 <MIL>
A:Cross-references: GB:M61862; NID:g211259; PIDN:AAA48627.1; PID:g211261

Query Match 22.5%; Score 85; DB 2; Length 321;
Best Local Similarity 26.8%; Pred. No. 3.7;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

Oy 7 HHHGSMAASGDLK-----NKVAQLKRVKSLKDKAAELKQVSRLENIEDL 52
Db 185 HLAERKDSTADLKLAAKLVEQREAVEERDSQLKQYKELSRATNLTKQLKLENIET 244
Oy 53 LKAKIGDLNNTSGIRPAKL 74
Db 245 VEKHLAKI-----GIRAPNLKLH 262

RESULT 7
F75103
conserved hypothetical protein PAB0812 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 02-Feb-2001
C:Accession: F75103
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: F75103
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-880 <KAM>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB50131.1; PID:g545866
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0812
C:Superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032

Query Match 22.5%; Score 85; DB 2; Length 880;
Best Local Similarity 37.2%; Pred. No. 10;
Matches 16; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

Oy 18 LKNKVAQLKRVKSLKDKAAELKQVSRLENIEDLAKIGDL 60
Db 240 IKGKISLKIQVEKLRKKRKGLEKIVQIERSTEEKAKISL 282

RESULT 8
JQ0894
P115 protein - Mycoplasma hyorhinis
C:Species: Mycoplasma hyorhinis
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 02-Feb-2001
C:Accession: JQ0894
R:McIntosh, M.A.; Wise, K.S.
Gene 97, 77-85, 1991
A:Title: A Mycoplasma hyorhinis protein with sequence similarities to nucleotide-binding
A:Reference number: JQ0894; MIMD:91138990
A:Accession: JQ0894
```

```
A:Molecule type: DNA
A:Residues: 1-979 <NOT>
A:Cross-references: GB:M34956
A:Note: the authors translated the codon AAA for residue 956 as Leu
C:Comment: This protein is located in the cytoplasm.
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: chromosome segregation protein SWC1
C:Keywords: nucleotide binding; P-loop
F:32-39/Region: nucleotide-binding motif A (P-loop)

Query Match 22.5%; Score 85; DB 2; Length 979;
Best Local Similarity 30.9%; Pred. No. 11;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

Oy 17 DLKKNVAQLKRVKSL-----KDKAAELKQVSRLENIEDLAKIGDLNNTSGI--- 66
Db 323 DQKTKIEIKQVSLKIQIQAASKQREIELDQUTRLNKAANSIKLQENDINKKIGVLE 382
Oy 67 RRPAAKL 74
Db 383 KKSAAAN 390

RESULT 9
F75216
hypothetical protein PAB2181 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F75216
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A:Reference number: A75001
A:Accession: F75216
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-281 <KAM>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49181.1; PID:g545
C:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2181

Query Match 22.4%; Score 84.5; DB 2; Length 281;
Best Local Similarity 40.8%; Pred. No. 3.6;
Matches 20; Conservative 10; Mismatches 12; Indels 7; Gaps 1;

Oy 17 DLKKNVAQLKRVKSL-----KVRSLKDKAAELKQVSRLENIEDLAKIG 58
Db 222 ELERKVSLESLNEVETKVKSLKKEKELENKVKLEEEVKNLKEGIG 270

RESULT 10
T29145
hypothetical protein F56A3.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29145
R:Geisler, C.; Bradshaw, H.; Wamsley, P.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid F56A3.
A:Reference number: Z20578
A:Accession: T29145
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1199 <GEI>
A:Cross-references: EMBL:U60446; PIDN:AAB37802.1; GSPDB:GN00019; CESP:F56A3.4
A:Experimental source: strain Bristol N2; clone F56A3
C:Genetics:
A:Gene: CESP:F56A3.4
A:Map position: 1
```


Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A83650; MUID:20263314
A:Accession: G83960
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 11188 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; MID:g10174886; PIDN:BA06206.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: smc

Query Match 21.8%; Score 82; DB 2; Length 1188;
Best Local Similarity 35.0%; Pred. No. 24;
Matches 21; Conservative 12; Mismatches 17; Indels 10; Gaps 1;

OY 11 GSMASGDLKHK-----VQLKRRVSLKDKAAELKQVSRLENIEDLKAKIGDL 60
DB 659 GSMGGSVKQKQPPPLGRQRELETKLQLEKAAKLEQEVKQKNERTELQQLIEDL 718

RESULT 15

H69402
hypothetical protein AF1225 - *Archaeoglobus fulgidus*
C:Species: *Archaeoglobus fulgidus*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: H69402
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artlisch, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343
A:Accession: H69402
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <KLE>
A:Cross-references: GB:AE001019; GB:AE000782; MID:g2689342; PIDN:AAB90025.1; PID:g264936

Query Match 21.6%; Score 81.5; DB 2; Length 212;
Best Local Similarity 31.1%; Pred. No. 4.7;
Matches 19; Conservative 15; Mismatches 20; Indels 7; Gaps 1;

OY 17 DLKRVKQQL-----KRVKSLKDKAAELKQVSRLENIEDLKAKIGDUNNTSGIRRP 69
DB 130 DLQKAKMLGIONELQKQIEELQKLERAKTELQKRRKSDLELEKIRELNROSSIYRL 189

OY 70 A 70
DB 190 A 190

Search completed: July 3, 2001, 14:59:47
Job time: 372 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:51 ; Search time 37.35 Seconds
(without alignments)
67.869 Million cell updates/sec

Title: US-09-490-291-11

Sequence: 1 MRGSHHHHSGMASGDLKN.....AKIGDLNNTSGTRPAAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	25.1	1186	1 SMC_BACSU	P11834 bacillus su
2	86.5	22.9	323	1 MAFI_RAT	P54842 rattus norv
3	86	22.8	548	1 CEAR_ECOLI	Q47502 escherichia
4	85.5	22.7	323	1 MAFI_MOUSE	P34841 mus musculu
5	85	22.5	979	1 P115_MYCHR	P15108 mycoplasma
6	82	21.8	766	1 STB6_YEAST	P26085 saccharomyc
7	81	21.5	166	1 YB38_MYCPN	P52860 mycoplasma
8	80	21.2	102	1 Y334_MERJA	O57780 methanococc
9	80	21.2	1225	1 SMC1_YEAST	P32908 saccharomyc
10	79.5	21.1	365	1 HEMA_IHAHL	P16994 influenza a
11	79.5	21.1	565	1 HEMA_IHAHO	P17000 influenza a
12	79	21.0	369	1 TMAF_AVISA	P23091 avian muscu
13	78.5	20.8	564	1 HEMA_IAGRE	P19698 influenza a
14	78.5	20.8	565	1 HEMA_IAMHI	P15658 influenza a
15	78	20.7	1938	1 MYS_AEOIR	P24723 aequipecten
16	78	20.7	4349	1 DYHC_FUSSO	P78716 fusarium so
17	77	20.4	461	1 US45_LACLC	P22865 lactococcus
18	76.5	20.3	741	1 BSG2_DROME	P11929 dirosophila
19	75.5	20.0	550	1 HEMA_IADHK	P13257 influenza a
20	75.5	20.0	1509	1 MYSN_ACACB	P05659 acanthamoeb
21	75	19.9	715	1 CLPB_MYCPN	P75247 mycoplasma
22	74.5	19.8	830	1 PAM1_YEAST	P7304 saccharomyc
23	74.5	19.6	879	1 MYSR_DROME	P35415 dirosophila
24	74	19.6	539	1 MYS3_HYDAT	P39922 hydra atten
25	74	19.6	1957	1 YD86_SCHPO	Q10411 schistosach
26	74	19.6	2022	1 ANTI_ONCVO	P21249 onchocerca
27	73.5	19.5	550	1 HEMA_IABAN	P03441 influenza a
28	73.5	19.5	550	1 HEMA_IADHI	P12532 influenza a
29	73.5	19.5	550	1 HEMA_IADH2	P12533 influenza a
30	73.5	19.5	550	1 HEMA_IADH3	P12534 influenza a
31	73.5	19.5	550	1 HEMA_IADH4	P12535 influenza a
32	73.5	19.5	550	1 HEMA_IADH6	P12557 influenza a
33	73.5	19.5	550	1 HEMA_IADH7	P12558 influenza a

ALIGNMENTS

RESULT	1	STANDARD	PRT	1186 AA.
SMC_BACSU				
ID	SMC_BACSU			
AC	P11834; 031735;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	CHROMOSOME PARTITION PROTEIN SMC.			
GN	SMC.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RA	Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;			
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 1171-1186 FROM N.A.			
RC	STRAIN=168;			
RX	MEDLINE=96093930; PubMed=7584053;			
RA	Oguro A., Kakeshita H., Honda K., Takamatsu H., Nakamura K.,			
RA	Yamane K.;			
RT	"src: a Bacillus subtilis gene encoding a homologue of the alpha-			
RT	subunit of the mammalian signal recognition particle receptor.";			
RL	DNA Res. 2:95-100(1995).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98367134; PubMed=9701812;			
RA	Moriya S., Tsujikawa E., Hassan A.K., Asai K., Kodama T.,			
RA	Ogasawara N.;			
RT	"A Bacillus subtilis gene-encoding protein homologous to eukaryotic			
RT	SMC motor protein is necessary for chromosome partition.";			
RL	Mol. Microbiol. 29:179-187(1998).			
RN	[5]			
RP	CHARACTERIZATION.			
RX	MEDLINE=98241483; PubMed=9573042;			
RA	Britton R.A., Lin D.C., Grossman A.D.;			
RT	"Characterization of a prokaryotic SMC protein involved in chromosome			
RT	partitioning.";			
RL	Genes Dev. 12:1254-1259(1998).			
CC	-1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND			
CC	PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION.			
CC	-1- SIMILARITY: BELONGS TO THE SMC FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: D64116; BAA10977.1; -

DR EMBL: 299112; CAB13467.1; -

DR EMBL: D49781; BAA08615.1; -

DR Subtilist; Bg11538; smc.

KW ATP-binding; Coiled coil.

FT NP_BIND 31 38

FT DOMAIN 244 481

FT 664 943

FT 989 1031

FT 50 50

FT 162 162

FT 175 175

FT 178 178

FT 192 192

FT 228 228

FT 236 236

FT 264 264

FT 271 271

FT 284 284

FT 309 309

FT 336 342

FT 365 365

FT 438 438

FT 444 444

FT 475 475

FT 494 494

FT 515 515

FT 542 542

FT 546 546

FT 586 600

FT CONFLICT 623 632

FT CONFLICT 664 664

FT CONFLICT 676 676

FT CONFLICT 680 680

FT CONFLICT 694 694

FT CONFLICT 701 701

FT CONFLICT 726 726

FT CONFLICT 738 740

FT CONFLICT 1186 AA; 135510 MW; 0163227AAFBACA2B CRC64;

Query Match

Best Local Similarity 25.1%; Score 94.5; DB 1; Length 1186;

Matches 25; Conservative 11; Mismatches 21; Indels 11; Gaps 2;

QY 11 GSMASDLKKNVAQLKRVSLKD-----KAELEKQEVSRLENIEDIKAKIGDL 60

DB 658 GSMTGAVKKNNKSLGRLEEDVTKRLAEMWEKTKALLEQEVKTKLSIQMEKTLADL 717

QY 61 NNT-SGIR 67

DB 718 RETGEGLR 725

RESULT 2

MAFL_RAT STANDARD; PRT; 323 AA.

ID MAFL_RAT

AC P54842;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DR 01-OCT-2000 (Rel. 40, Last annotation update)

DE TRANSCRIPTION FACTOR MAFL.

GN MAFB OR MAFL.

OS Rattus norvegicus (Rat).

CC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WISTAR; TISSUE=Liver;

RA MEDLINE=97190228; PubMed=9038383;

RA Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,

RA Kuboki Y., Nishizawa M., Nishi S.;

RT "Rat maf related genes: specific expression in chondrocytes, lens and

RT spinal cord.";

RL Oncogene 14:745-750(1997).

CC -1- SUBCELLULAR LOCATION: NUCLEAR.

CC -1- SIMILARITY: BELONGS TO THE B2IP FAMILY. MAF SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U56241; AB50062.1; -

DR HSSP; P05412; IJUN.

KW Transcription regulation; DNA-binding; Nuclear protein.

FT DOMAIN 131 143

FT 158 167

FT 194 201

FT 238 264

FT 266 287

FT 323 AA; 35792 MW; 6E386340D1F840A5 CRC64;

SO SEQUENCE

Query Match

Best Local Similarity 22.9%; Score 86.5; DB 1; Length 323;

Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

QY 5 HHHHH-----GSMAS----- 15

DB 162 HHHHHQASPPSSAASPAQQLPTSHPGPHAAATAAGSNGVEDRFSDQLVMSV 221

QY 16 -----GDLKNKYAQLKRVSLKD-----AAELK 40

DB 222 RELNHLRGFTDEVIKRLKOKRRRTLKNRYAOSCRKYRVOQKHLENEKTOLOQVEDLK 281

QY 41 QEVSRLENIEDIKAKIGDLNNTSGIR 67

DB 282 QEVSRLENIEDIKAKIGDLNNTSGIR 307

RESULT 3

CEAK_ECOLI STANDARD; PRT; 548 AA.

ID CEAK_ECOLI

AC Q47502; P75615;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DR 01-OCT-2000 (Rel. 40, Last annotation update)

DE COLICIN K.

GN CKA.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K49;

RX MEDLINE=96074330; PubMed=7592493;

RA Pliel H., Braun V.;

RT "Strong function-related homology between the pore-forming colicins K

RT and S.";

RL J. Bacteriol. 177:6973-6977(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA Izard J., Chartier M., Baty D.;

RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS COLICIN IS A CHANNEL-FORMING COLICIN. THIS CLASS OF


```

CC TRANSMEMBRANE TOXINS DEPOLARIZE THE CYTOPLASMIC MEMBRANE, LEADING
CC TO DISSIPATION OF CELLULAR ENERGY.
CC -1- FUNCTION: COLICINS ARE POLYPEPTIDE TOXINS PRODUCED BY AND ACTIVE
CC AGAINST, ESCHERICHIA COLI AND CLOSELY RELATED BACTERIA.
CC -1- SIMILARITY: BELONGS TO THE CHANNEL FORMING COLICIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X87834; CA61099.1; -
CC EMBL: U27452; AAB41288.1; -
CC InterPro: IPR000293; -
CC Pfam: PF01024; Colicin; 1.
CC PRINTS: PR00280; CHANCOLICIN.
CC PROSITE: PS00276; CHANNEL_COLICIN; 1.
CC Antidiotic; Bacteriocin; Plasmid; Transmembrane.
CC TRANSMEM 505 525 POTENTIAL.
CC CONFLICT 155 155 S->G (IN REF. 2).
CC SEQUENCE 548 AA; 59661 MW; 2E67689D088CFC86 CRC64;
SO
Query Match 22.8%; Score 86; DB 1; Length 548;
Best Local Similarity 32.1%; Pred. No. 1.8;
Matches 18; Conservative 13; Mismatches 15; Indels 10; Gaps 1;
OY 6 HHHHGSMAAGDLKKNVAOLKRRYSLKDKAELKOEVSRLNTEEDLAKIGLGN 61
DB 188 HSHYHG-----ITDLERVDYDLQKRSNLDADISKLSYKNTLQSGIDVN 233
RESULT 4
MAP1_MOUSE STANDARD; PRT; 323 AA.
AC P59841;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTION FACTOR MAP1 (SEGMENTATION PROTEIN KR) (KREISLER).
GN MAP1 OR MAP1 OR KRM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95094266; Pubmed=8001130;
RA Cordes S.P., Barsh G.S.;
RT "The mouse segmentation gene kr encodes a novel basic domain-leucine
RT zipper transcription factor.";
RL Cell 79:1025-1034(1994).
CC -1- FUNCTION: MAY PLAY AN EARLY ROLE IN AXIAL PATTERNING (HINDBRAIN
CC SEGMENTATION).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, GUT, LUNG, AND BRAIN.
CC -1- DEVELOPMENTAL STAGE: DETECTABLE AT 8.0 DPC (ONE SOMITE) AS A BAND
CC IN THE CAUDAL HINDBRAIN, AND BY 8.5 DPC (SIX TO EIGHT SOMITES),
CC THE HIGH LEVEL DOMAIN EXHIBITS A SHARP ROSTRAL EDGE COINCIDENT
CC WITH THE R4/R5 BOUNDARY AND A DIFFUSE CAUDAL EDGE LOCATED MIDWAY
CC THROUGH R6.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: L36435; AAA65689.1; -
CC DR HSSP; P05412; IJUN.
CC DR TRANSFAC; T01439; -
CC DR MGD; MGI:104555; MafB.
CC KW Transcription regulation; DNA-binding; Nuclear protein.
CC FT DOMAIN 131 143 POLY-HIS.
CC FT DOMAIN 158 167 POLY-HIS.
CC FT DNA_BIND 238 264 BASIC MOTIF.
CC FT DOMAIN 266 287 LEUCINE-ZIPPER.
CC FT NOTAGEN 248 248 N->S: LOSS OF TRANSCRIPTIONAL ACTIVITY.
CC SEQUENCE 323 AA; 35809 MW; D77AE07ABD9C2AD2 CRC64;
SO
Query Match 22.7%; Score 85.5; DB 1; Length 323;
Best Local Similarity 23.1%; Pred. No. 1.1;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;
OY 5 HHHHH-----GSMAS-----
DB 162 HHHHHHQAAPPSSAAPQOLPTSHQPGPHATTAAGNGSVEDSDQLVMSV 221
OY 16 -----GDLKNRYAOLKRRYSLKDK-----AAELK 40
DB 222 RELNRHLRGFTKDEVIRLKQKRTLRKNGYAGSCRYKRVQKHLLENKTLIQVQLK 281
OY 41 QEVSRLENEDLAKIGDNLNTSGIR 67
DB 282 QEVSLARENDAYKVECKLAN-SGFR 307
RESULT 5
P115_MYCHR STANDARD; PRT; 979 AA.
AC P41508;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE P115 PROTEIN.
OS Mycoplasma hyorhinis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91138990; Pubmed=1825306;
RA Notariicola S.M., McIntosh M.A., Wise K.S.;
RT "A Mycoplasma hyorhinis protein with sequence similarities to
RT nucleotide-binding enzymes.";
RL Gene 97:77-85(1991).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. SIMILAR TO OTHER MYCOPLASMA
CC P115.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M34956; AAA25423.1; -
CC DR PIR; JQ0894; JQ0894.
CC KW ATP-binding; Coiled coil.
CC FT NP_BIND 32 39
CC FT DOMAIN 169 224 COILED COIL (POTENTIAL).
CC FT DOMAIN 231 400 COILED COIL (POTENTIAL).
CC FT DOMAIN 569 821 COILED COIL (POTENTIAL).
CC FT DOMAIN 884 912 ALA/ASP-RICH (DA-BOX).
CC SEQUENCE 979 AA; 110566 MW; 30D51C56B56280F4 CRC64;

```

```

Query Match          22.5%; Score 85; DB 1; Length 979;
Best Local Similarity 30.9%; Pred. No. 4;
Matches 21; Conservative 15; Mismatches 22; Indels 10; Gaps 2;

QY 17 LKKNVAQLKRRVRSI-----KDKAELKQEVSRLENEIDLAKIDLNTSGI---- 66
DB 323 DQKTKIEEIKQVESIKIQINASKOREIELDQOLTRLNKANSIKLQENDINKEIGVLE 382

QY 67 RRPAAKLN 74
DB 383 KKSAAAN 390

RESULT 6
STB6_YEAST
ID STB6_YEAST STANDARD; PRT; 766 AA.
AC P36085;
DT 01-JUN-1994 (Rel. 29, Created)
DR 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE STB6 PROTEIN
GN STB6 OR YKL072W OR YKL352.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins."
RT yeast 10:569-574(1994).
RN [2]
RP SEQUENCE OF 1-557 FROM N.A.
RA Pohl T.M., Pohl F.M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98055153; PubMed=9393435;
RA Kaestlen M.M., Stillman D.J.;
RT "Identification of the Saccharomyces cerevisiae genes STB1-STB5
RT encoding Sin3p binding proteins."
RL Mol. Gen. Genet. 256:376-386(1997).
CC -1- FUNCTION: BINDS TO SIN3.
CC -1- SIMILARITY: STRONG, TO YEAST STB2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC
CC EMBL; X75780; CAA53402.1; -
DR EMBL; 228072; CAA81909.1; -
DR PIR; S37894; S37894.
DR PIR; S39170; S39170.
DR PIR; S44515; S44515.
DR SGD; S000155; YKL072W.
SQ SEQUENCE 766 AA; 88835 MW; 9BCBEA2EE03A9AF1 CRC64;

Query Match          21.8%; Score 82; DB 1; Length 766;
Best Local Similarity 26.7%; Pred. No. 5;
Matches 20; Conservative 20; Mismatches 25; Indels 10; Gaps 2;

```

```

QY 2 RGSHHH-HHGSMA-----GDLKNVAQLKRRVRSIKDKAELKQEVSRLENEIE 51
DB 620 RGSNEHFWEYGNMNOQLIESEKKDKMSLSSKLKYNVILDRRVADVBSVDHFDRKLE 679

QY 52 DLKATIGDLNTSGI 66
DB 680 DVRKILLEQNNSKDI 694

RESULT 7
YB38_MYCPN
ID YB38_MYCPN STANDARD; PRT; 166 AA.
AC P75260;
DT 01-OCT-2000 (Rel. 40, Created)
DR 01-OCT-2000 (Rel. 40, Last sequence update)
DE 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MPN138 (E07_ORF166).
GN MPN138 OR MP016.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0134 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC
CC EMBL; AE000003; AAB95664.1; -
DR InterPro; IPR002862; -
DR Pfam; PF01519; DUF16; 1.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 19518 MW; BE44F5377B2FA709 CRC64;

Query Match          21.5%; Score 81; DB 1; Length 166;
Best Local Similarity 37.2%; Pred. No. 1;
Matches 16; Conservative 15; Mismatches 12; Indels 0; Gaps 0;

QY 18 LKKNVAQLKRRVRSIKDKAELKQEVSRLENEIDLAKITGL 60
DB 81 IENKVDKLEVKDKLEKVKLEAKYDKLEKVKLEAKYDKL 123

RESULT 8
Y334_METJA
ID Y334_METJA STANDARD; PRT; 102 AA.
AC Q57780;
DT 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0334.
GN MJ0334.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.

```


DR EMBL; M24721; AAA43100.1; ALT_SEQ.
DR PIR; D34064; HMYE4.
DR HSSP; P03437; 1HTM.
DR InterPro; IPR000149; -
DR InterPro; IPR000386; -
DR InterPro; IPR001364; -
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 16 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 17 340 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 345 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63831 MW; BA533050DC3F186B CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 6.5;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

OY 3 GSHHHHHSMASGDLKNVA-----OLKRRVSLKKAELKOEVRLENEIEDLKA 55
DB 367 GRRHNSGEGAGDLKSTQAALIDQINGKLNRYIEKTKNEKHFQIEKSEFVGRIDLEK 426

OY 56 KIGD 59
DB 427 YVED 430

RESULT 11
HEMA_IAMTO STANDARD; PRT; 565 AA.
ID HEMA_IAMTO
AC P17000; 084002; 084003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HAI CHAIN;
HEMAGGLUTININ HAZ CHAIN].
GN HA
OS Influenza A virus (strain A/Equine/Tokyo/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11418;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8920489; PubMed=2705299;
RA Kawoka Y., Bean W.J., Webster R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RT Virology 169:283-292(1989)
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24720; AAA43111.1; ALT_SEQ.

DR PIR; C34064; HMYE3.
DR HSSP; P03437; 1HTM.
DR InterPro; IPR000149; -
DR InterPro; IPR000386; -
DR InterPro; IPR001364; -
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
DR Envelope protein; Hemagglutinin; Glycoprotein; Signal.
KW SIGNAL
FT CHAIN 1 16 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 17 343 HEMAGGLUTININ HAZ CHAIN.
FT CARBOHYD 345 565 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 565 AA; 63580 MW; 84B7D4AD70629B7A CRC64;

Query Match 21.1%; Score 79.5; DB 1; Length 565;
Best Local Similarity 28.1%; Pred. No. 6.5;
Matches 18; Conservative 15; Mismatches 24; Indels 7; Gaps 1;

OY 3 GSHHHHHSMASGDLKNVA-----OLKRRVSLKKAELKOEVRLENEIEDLKA 55
DB 367 GRRHNSGEGAGDLKSTQAALIDQINGKLNRYIEKTKNEKHFQIEKSEFVGRIDLEK 426

OY 56 KIGD 59
DB 427 YVED 430

RESULT 12
TMAF_AVIS4 STANDARD; PRT; 369 AA.
ID TMAF_AVIS4
AC P23091;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSFORMING PROTEIN TMAF.
GN V-MAF.
OS Avian musculoponeurotic fibrosarcoma virus AS42.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11873;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9004665; PubMed=2554284;
RA Nishizawa M., Kataoka K., Goto N., Fujiwara K.T., Kawai S.;
RT "v-maf, a viral oncogene that encodes a 'leucine zipper' motif.";
RT Proc. Natl. Acad. Sci. U.S.A. 86:7711-7715(1989).
CC -1- FUNCTION: MIGHT BE A TRANSCRIPTIONAL TRANS-ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: INDICES MUSCULOAPONEUROTIC FIBROSARCOMA IN CHICKENS.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A ENV-MAF
POLYPROTEIN.
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M26769; AAA42377.1; -
DR PIR; B33975; TVFAF.
DR HSSP; P05412; 1JUN.

```

DR TRANSFAC: T01430:
KW Transforming protein; Transcription regulation; DNA-binding;
FT Nucleic protein.
FT DOMAIN 163 167 POLY-ALA.
FT DOMAIN 173 182 HIS-RICH.
FT DOMAIN 184 191 POLY-GLY.
FT DOMAIN 214 223 POLY-GLY.
FT DOMAIN 233 240 POLY-GLY.
FT DNA_BIND 274 300 BASIC MOTIF.
FT DOMAIN 302 323 LEUCINE-ZIPPER.
SQ SEQUENCE 369 AA: 38892 MW: F386B220ACE50FF6 CRC64:

Query Match 21.0%; Score 79; DB 1; Length 369;
Best Local Similarity 24.1%; Pred. No. 4.6;
Matches 26; Conservative 8; Mismatches 20; Indels 54; Gaps 2;

OY 3 GSHHHHSGMASGDL-----KKNVAOLKRRVSLDK 35
DB 226 GLNHPHGGGGGGGLHFDREDFSDQLVTMSRELNLQGLVSGKEEYIRLKKRRTLKNR 285
OY 36 -----AELKQVSRLENEIEDLAK 56
DB 286 GYAGSCRFKRVQRRHYLESEKNQLDOVEHLKQELSRVREDAVYKREK 333

RESULT 13
HEMA_IAGRE
ID HEMA_IAGRE STANDARD; PRT; 564 AA.
AC P19698;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Grey teal/Australia/2/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A virus.
OC Influenza A virus.
OX NCBL_TaxID=11385;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Donls R.O., Bean W.J., Kawaka Y., Webster R.G.;
RT "Distinct lineages of influenza virus H4 hemagglutinin genes in
RT different regions of the world.";
RL Virology 169:408-417(1989).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M25284; AAA43217.1; -.
CC PIR: B34214; HMIVE2.
CC HSSP: P03437; 2HMG.
CC InterPro: IPR000149; -.
CC InterPro: IPR000386; -.
CC InterPro: IPR001364; -.
CC Pfam: PF00509; Hemagglutinin; 1.
CC PRINTS: PR00329; HEMAGGLUTIN1.
CC PRINTS: PR00330; HEMAGGLUTIN1.
CC PRINTS: PR00331; HEMAGGLUTIN2.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.

```

```

FT SIGNAL 1 16
FT CHAIN 17 342 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 344 564 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 14 14 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 34 34 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 564 AA: 63237 MW: A4CD347146D0971C CRC64:

Query Match 20.8%; Score 78.5; DB 1; Length 564;
Best Local Similarity 28.1%; Pred. No. 7.9;
Matches 18; Conservative 14; Mismatches 25; Indels 7; Gaps 1;

OY 3 GSHHHHSGMASGDLKKNVA-----QLKRRVSLDKRAELKQVSRLENEIEDLKA 55
DB 366 GFRHQNMEGTGTADLTSTQALDQINGKLNRLEIKNDKHQIEKEFEYEGRIQLK 425
OY 56 KIGD 59
DB 426 YVED 429

RESULT 14
HEMA_IAHMI
ID HEMA_IAHMI STANDARD; PRT; 565 AA.
AC P15658; 083988; 083987; 067097; 067098;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Miami/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A virus.
OC Influenza A virus.
OX NCBL_TaxID=11407;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204899; PubMed=3973560;
RA Daniels R.S., Skelhel J.J., Willey D.C.;
RT "Amino acid sequences of haemagglutinins of influenza viruses of the
RT H3 subtype isolated from horses.";
RL Virology 169:283-292(1989).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M24719; AAA43105.1; ALT_SEQ.
CC EMBL: M29257; AAA43164.1; -.
CC PIR: B34064; HMIVE2.
CC HSSP: P03437; 1HTM.
CC InterPro: IPR000149; -.

```

```

DR InterPro: IPR000386; -.
DR InterPro: IPR001364; -.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PRO0329; HEMAGGLUTN12.
DR PRINTS: PRO0330; HEMAGGLUTN1.
DR PRINTS: PRO0331; HEMAGGLUTN2.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 37 37 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 180 180 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 13 15 WNH -> AAD (IN REF. 2).
FT CONFLICT 73 73 RV -> SG (IN REF. 2).
FT CONFLICT 173 173 E -> G (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 377 377 L -> G (IN REF. 2).
FT CONFLICT 380 380 G -> A (IN REF. 2).
FT CONFLICT 400 400 I -> F (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63729 MW; EA1FEEDD40AC2A CRC64;

Query Match 20.8%; Score 78.5; DB 1; Length 565;
Best Local Similarity 26.6%; Pred. No. 7.9;
Matches 17; Conservative 17; Mismatches 23; Indels 7; Gaps 1;

QY 3 GSHHHHNGSMAGDLKKNVA-----QLKRVKSLDKAAELKQVSRLENEIEDLKA 55
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 367 GPRVNSGCTIQAGDLKSTQALIDQINKLRVIEKTEKHFQIKESSEVEGRIDLEK 426
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 56 KIGD 59
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 427 YVED 430
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 15
MS_ABOIR STANDARD; PRT; 1938 AA.
ID MS_ABOIR
AC P24733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN, STRIATED MUSCLE.
OS Aequipecten irradians (Bay scallop).
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;
OC Pectinoidea; Pectinidae; Argopecten.
OX NCBI_TaxID=31199;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RA MEDLINE=92011595; PubMed=1917970;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Complete primary structure of a scallop striated muscle myosin heavy
RT chain. Sequence comparison with other heavy chains reveals regions
RT that might be critical for regulation.";
RL J. Biol. Chem. 266:18469-18476(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Adductor muscle;
RA MEDLINE=91088319; PubMed=2263488;
RA Nyitray L., Goodwin E.B., Szent-Gyorgyi A.G.;
RT "Nucleotide sequence of full length cDNA for a scallop striated
RT muscle myosin heavy chain.";
RL Nucleic Acids Res. 18:7158-7158(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 777-836.
RA MEDLINE=94173332; PubMed=8127365;

```

```

RA Xie X., Harrison D.H., Schlichting I., Sweet R.M., Kalabokis V.N.,
RA Szent-Gyorgyi A.G., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2.8-A
RT resolution.";
RL Nature 368:306-312(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RA MEDLINE=96419133; PubMed=8605510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
RT resolution: Implications for regulation.";
RL Structure 4:21-32(1996).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X55714; CAA39247.1; -.
DR PIR: S13557; S13557.
DR PIR: A40997; A40997.
DR PDB: 1SCM; 30-APR-94.
DR PDB: 1WDC; 11-JUL-96.
DR InterPro: IPR000048; -.
DR InterPro: IPR001609; -.
DR InterPro: IPR002928; -.
DR Pfam: PF00612; IQ; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF00063; myosin_head; 1.
DR PRINTS: PRO0193; MYOSINHEAVY.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Alkylation; Heptad repeat pattern; 3D-structure.
FT DOMAIN 1 835 GLOBULAR HEAD (S1).
FT DOMAIN 836 1938 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 836 1938 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 778 805 COILED COIL (POTENTIAL).
FT NP_BIND 176 183 IQ.
FT MOD_RES 693 693 ATP (BY SIMILARITY).
FT MOD_RES 703 703 ALKYLATION (SH-1) (BY SIMILARITY).
FT HELIX 778 821 ALKYLATION (SH-2) (BY SIMILARITY).
FT TURN 822 823
FT HELIX 825 833
FT TURN 834 835
SQ SEQUENCE 1938 AA; 222821 MW; A5CCEA127D1A4896 CRC64;

Query Match 20.7%; Score 78; DB 1; Length 1938;
Best Local Similarity 31.4%; Pred. No. 32;
Matches 16; Conservative 15; Mismatches 20; Indels 0; Gaps 0;

QY 13 MASGLKKNVAKLRKRVSLDKAAELKQVSRLENEIEDLKAIGDIANT 63
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 907 MQKADPESQIKELERLDEEDAAADLEGIKKKMADNANLKKDIGDLENT 957
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

Search completed: July 3, 2001, 15:10:52
 Job time: 752 sec

Fri Jul 6 08:02:47 2001

us-09-490-291-11.rsp

Page 9

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:42 ; Search time 95:41 Seconds
(without alignments)
102,616 Million cell updates/sec

Title: US-09-490-291-11

Sequence: 1 MRGSHHHHHGSMASGLKN.....AKIGDNNNTSGIRPAKLN 74

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTOZOA:*
12: SP_UNCLASSIFIED:*
13: SP_VERTEBRATE:*
14: SP_VIRUS:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	24.1	481	2	09LAX5
2	89.5	23.7	356	13	073679
3	88.5	23.5	146	13	091891
4	88.5	23.5	286	13	057342
5	88	23.3	259	2	066901
6	87	23.1	324	13	09W6B1
7	86	22.8	155	7	031405
8	85.5	22.7	286	13	042290
9	85.5	22.7	323	4	09H1F1
10	85	22.5	321	7	031620
11	85	22.5	479	2	09LAX2
12	85	22.5	480	2	09LAX3
13	85	22.5	880	1	09U7C8
14	84.5	22.4	281	1	09V217
15	84.5	22.4	323	4	09V5Q3
16	84	22.3	1199	5	P91349
17	82	21.8	173	5	001354
18	82	21.8	311	13	090888
19	82	21.8	313	13	09P0A6

20	82	21.8	461	2	09LAX6	09LAX6 streptococ
21	82	21.8	1188	2	09KA07	09KA07 bacillus ha
22	81.5	21.6	212	1	029043	029043 archaeoglob
23	81.5	21.6	240	9	09MC79	09MC79 bacterioph
24	81.5	21.6	2244	5	09NCG0	09NCG0 drosophila
25	81	21.5	270	1	058907	058907 pyrococcus
26	81	21.5	396	3	09Y7S2	09Y7S2 schistosach
27	81	21.5	409	1	09YAC7	09YAC7 aeropyrum p
28	81	21.5	528	5	09TY57	09TY57 schistosoma
29	81	21.5	802	5	096398	096398 schistosoma
30	81	21.5	1940	5	09U7E3	09U7E3 pecten maxi
31	80.5	21.4	269	2	054615	054615 legionella
32	80.5	21.4	647	1	09H0M8	09H0M8 halobacteri
33	80	21.2	311	13	090370	090370 coturnix co
34	80	21.2	528	5	026589	026589 schistosoma
35	80	21.2	1624	5	09U679	09U679 strongyloce
36	80	21.2	1940	5	002456	002456 schistosoma
37	79.5	21.1	566	14	091GSO	091GSO influenza a
38	79	21.0	318	5	09LXR8	09LXR8 arabidopsis
39	79	21.0	353	5	09VUK4	09VUK4 drosophila
40	79	21.0	369	13	092171	092171 gallus gall
41	79	21.0	407	2	09WZHO	09WZHO thermotoga
42	79	21.0	426	2	09LAX5	09LAX5 streptococ
43	79	21.0	554	2	09WZ38	09WZ38 thermotoga
44	78.5	20.8	1956	5	020641	020641 caenorhabdi
45	78	20.7	442	5	09V7N6	09V7N6 drosophila

ALIGNMENTS

RESULT	ID	09LAX5	PRELIMINARY:	PRT:	481 AA.
DT	01-OCN-2000	(TREMblrel. 15, Created)			
DT	01-OCN-2000	(TREMblrel. 15, last sequence update)			
DT	01-MAR-2001	(TREMblrel. 16, last annotation update)			
DE	PSPA (FRAGMENT).				
GN	PSPA.				
OS	Streptococcus pneumoniae.				
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID-1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	SPRAIN-BG11703;				
RA	Hollingshead S.K., Becker R., Briles D.E.;				
RT	"Diversity of Pspa: mosaic genes and evidence for past recombination				
RT	in Streptococcus pneumoniae."				
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF071821; AAF27716.1; -				
DR	InterPro: IPR002965; -				
DR	PRINTS: PR01217; PRICHTEXTNSN.				
FT	NON_TER	481			
FT	SEQUENCE	481 AA; 53500 MW; EA3C66445EFCCE2B CRC64;			
QY	Query Match	24.1% Score 91; DB 2; Length 481;			
Db	Best Local Similarity	37.2% Pred. No. 3.1; Indels			
	Matches 16; Conservative 15; Mismatches 12; Gaps 0;				
QY	17	DLKKNVQLKRRVSLKDKAELEKQVSRLENEIEDLKAKIGD 59			
Db	215	ELQKKVADLEKEIADVKKTVADEKEVAKLEKDEGFESGDE 257			
RESULT	2				
073679					
ID	073679	PRELIMINARY;	PRT;	356 AA.	
DT	01-AUG-1998	(TREMblrel. 07, Created)			
DT	01-AUG-1998	(TREMblrel. 07, last sequence update)			

DT 01-MAR-2001 (TREMBLE). 16, last annotation update)
DE TRANSCRIPTION FACTOR VAL.
CN VALENTINO.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariopomys;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98165393; PubMed=9425134;
RA Moens C.B., Cordes S.P., Giorgioanni M.W., Barsh G.S., Kimmel C.B.;
RT "Equivalence in the genetic control of hindbrain segmentation in fish
RT and mouse.";
RL Development 125:381-391(1998).
DR EMBL: AF006641; AAC18821.1; -.
DR ZFIN: ZDB-GENE-980526-515; val.
DR InterPro: IPR001871; -.
DR SMART: SM00338; BRIL; 1.
SO SEQUENCE 356 AA; 40243 MW; 07420DB0F6CDD08F1 CRC64;

Query Match	23.7%	Score 89.5;	DB 13;	Length 356;
Best Local Similarity	22.2%;	Pred. No. 3.1;		
Matches 28; Conservative	13;	Mismatches	22;	Indels 63; Gaps 3;

Oy	5	HHHHH-----GSM-----	-AGSDLKAKYAOILRK	28
			: : : : :	
Db	215	HHHHHHRHGRGQGGHHGGGLNVEDRFSDDQLVTSMVSVELNHLGFLTKDEYIRLKK		274
Oy	29	VRSLKDK-----	-AALKEVSRLENEIDILKAKIGDIN	61
		: : : :	: : : :	
Db	275	RRLKNGGYAOSCFKFKYQAKHLENEKTLQINLOVEQLKDELINLRARERDYKILCKEKL		334
Oy	62	NTSGIR		67
		:		
Db	335	GANGFR		340

RESULT	3	
Q91891		
ID	Q91891	PRELIMINARY;
		PRT;
		146 AA

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, last annotation update)
DE BZIP TRANSCRIPTION FACTOR L-MAF (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-BREED FANOUJI, AND BREED G-BI LEGHORN; TISSUE=SPLEEN;
RA Zhou H., Lamont S.J.;
RT "Genetic variation among chicken lines and mammalian species in
RT specific genes."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF221558; AAF71240.1; -
DR EMBL; AF211557; AAF71239.1; -
DR InterPro: IPR001871; -
DR SMART: SM00338; BRLZ: 1.
FT NON_TER 1
FT NON_TER 146
FT NON_TER 146
SQ SEQUENCE 146 AA; 17579 MW; 382BADID098E1DDC CRC64;

```

4 SHHHHHH-----SMASGD-----KNKAQLRKRVSLKDK----- 35
Query Match      23.5%  Score 88.5;  DB 13;  Length 146;
Best Local Similarity 27.5%  Pred. No. 1.6;
Matches      20;  Conservative  11;  Mismatches  14;  Gaps  3;

```

[illegible]

RESULT	4	
057342		
ID	057342	PRELIMINARY;
		PRT;
		286 AA

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE BZIP TRANSCRIPTION FACTOR MAFa.

0S *Coturnix coturnix japonica* (Japanese quail).
0C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
0C Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
0C *Coturnix*.

RN [1]
RP SEQUENCE FROM N.A.
RA Benhelfa S., Felder-Schmittbuhl M.P., Calothy G.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases

```
DR InterPro; IPR001871; -
DR SMART; SM00338; BRLZ; 1.
SQ SEQUENCE 286 AA; 32463 MW; 0D4ED7BF08D91397 CRC64;
```

Query Match	23.5%	Score 88.5;	DB 13;	Length 286;
Best Local Similarity	27.5%	Pred. No. 3;		
Matches 28; Conservative	11;	Mismatches	14;	Indels 49; Gaps 3;

```

OY  4 SHHHHHHG-----SMASGL-----KKKVAQLKKRVSLKDK----- 35
      :|||||      ||:  :|      | :|||:| :|||:
Db 157 ANHHHHHHHDLREERSDDLVSMSVANELNRQLRGFSKEVEYIRLKKRRTLKNRGVAQSC 216

```

```

Qy 36 -----AAELKQEVSRLENEIEDLKAK 56
      :| | | | | | | : | |
Db 217 RYKRVQQRHILENEKCOLQSQVEQLKQEVSRLLAKERDLYKEK 258

```

RESULT	5
066901	
ID	066901
PRELIMINARY;	
PRT;	259 AA

DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HYDROGENASE EXPRESSION/FORMATION PROTEIN B.

OS Aquifex aeolicus;
OC Bacteria; Aquificales; Aquificaceae; Aquifex
OX NCBI_TaxID=63363;
RN [1]

RC SFRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Tenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,

RT "The complete genome of the hyperthermophilic bacterium *Agitex*
RT *aeolicus*,"
RL Number 392:353-358(1998).
EMBL; AEO00701; AAC06860.1; -.
DR EMBL; AEO00701; AAC06860.1; -.

DR InterPro: IPR002894; -.
DR Pfam: PF01495; HyPB_Ureg; 1.
DR PROSITE: PS00028; 2INC_FINGER_C2H2; UNKNOWN 1.
SQ SEQUENCE 259 AA; 29015 MW; 4522AC8CFA53806C CRC64;

Query Match 23.38; Score 88; DB 2; Length 259;
 Best Local Similarity 28.68; Pred. No. 3;
 Matches 26; Conservative 12; Mismatches 29; Indels 24; Gaps 3;

QY 5 HHHHNSMSGDLKNVAQLKRVSLKDKAAELKQ-----EVS 44
 DB 19 HHHHHTNPALGD--KRTVEYLKRLISANDEQAESNRHFERHIGILAVNIMSPGSGKTT 76
 QY 45 RLENEIEDLK-AKIGDLNMTSGIRPAK 73
 DB 77 LLERTIELLKDLELKIGYIEGDDLETRDAERI 107

RESULT 6
 Q9M6B1 PRELIMINARY; PRT; 324 AA.
 AC Q9M6B1; PRELIMINARY; PRT; 324 AA.
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE BASIC DOMAIN LEUCINE ZIPPER TRANSCRIPTION FACTOR.
 GN MAF.
 OS Brachydanio rerio (zebrafish) (zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 NC NCBL_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9173794; PubMed=10072793;
 RA Schwarzeisen M., Kirn A., Halfter P., Cordes S.P.;
 RT "Expression of Zkml2, a homologue of the Kml1/yai segmentation gene,
 RT during embryonic patterning of the zebrafish (Danio rerio).";
 RL Mech. Dev. 80:223-226(1999).
 DR EMBL: AF109781; AAD26141.1; -.
 DR InterPro: IPR001871; -.
 DR SMART: SM00338; BRIZ; 1.
 SO SEQUENCE 324 AA; 36025 MW; C1ED46858C011EAB CRC64;

Query Match 23.18; Score 87; DB 13; Length 324;
 Best Local Similarity 24.28; Pred. No. 4.5;
 Matches 32; Conservative 11; Mismatches 19; Indels 70; Gaps 4;

QY 5 HHHHNG-----SMASGDL-----KN 20
 DB 176 HHHHHGHOOHOTPASRPSSTNSLSPITROPDDRSDDQLVTSVRELNRHVRGVSK 235
 QY 21 KVAQLKRVSLKDKA-----AELKQVSRLENEIEDLK 55
 DB 236 EVILKQKRRRLKRGVAAOSSRYKRVQOHRHLEKTLQIHDHKLQKELSRVLRERDAYKE 295
 QY 56 KIGDLNMTSGIR 67
 DB 296 KYERLVN-SGFR 306

RESULT 7
 Q31405 PRELIMINARY; PRT; 155 AA.
 AC Q31405; PRELIMINARY; PRT; 155 AA.
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE MHC B-G ANTIGEN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBL_TaxID=9031;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=90077532; PubMed=2592020;
 RA Kaufman J., Salomonsen J., Skjodt K.;
 RT "B-G CDNA clones have multiple small repeats and hybridize to both
 RT chicken MHC regions.";
 RL Immunogenetics 30:440-451(1989).
 DR EMBL: M27668; AAA48624.1; -.
 KW MHC.
 FT NON_TER
 SO SEQUENCE 155 AA; 18201 MW; 7FC66EF2584F3978 CRC64;

Query Match 22.88; Score 86; DB 7; Length 155;
 Best Local Similarity 27.58; Pred. No. 2.7;
 Matches 22; Conservative 17; Mismatches 23; Indels 18; Gaps 2;

QY 7 HHHHNSMSGDLKNKVA-----OLKRVSLKDKAAELKQVSRLENEIED 52
 DB 19 HCNEMDLSADLKQIAELVQREGVERMHIQLRKHYEKLSRAANLTKQKLENEIEE 78
 QY 53 LKAKIGDLNMTSGIRPAK 72
 DB 79 VEKHL---KKTGRAPPLK 94

RESULT 8
 Q42290 PRELIMINARY; PRT; 286 AA.
 AC Q42290; PRELIMINARY; PRT; 286 AA.
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE BZIP TRANSCRIPTION FACTOR L-MAF.
 GN L-MAF.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCBL_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=LENS.
 RX MEDLINE=98192815; PubMed=9525857;
 RA Ogino H., Yasuda K.;
 RT "Induction of lens differentiation by activation of a bZIP
 RT transcription factor, L-Maf.";
 RL Science 280:115-118(1998).
 DR EMBL: AF034570; AAC15781.1; -.
 DR InterPro: IPR001871; -.
 DR SMART: SM00338; BRIZ; 1.
 SO SEQUENCE 286 AA; 32445 MW; 15DBD7B9B8DFA522 CRC64;

Query Match 22.78; Score 85.5; DB 13; Length 286;
 Best Local Similarity 25.98; Pred. No. 5.4;
 Matches 30; Conservative 12; Mismatches 25; Indels 49; Gaps 3;

QY 4 SHHHHNG-----SMASGDL-----KNKYAQLKRVSLKDK----- 35
 DB 157 AHHHHHHHLLEERFSDQLVSMYSVRLNQLGFSKEEYIRLQKONRTLKNNGYASOC 216
 QY 36 -----AELKQVSRLENEIEDLKAKIGDLNMTSGIRPA 70
 DB 217 RYKRVQOHRHLENEKQLOSOVQDLKQVSRLEAKERDLYKKEYEKLARGPPEPS 272

RESULT 9
 Q9H1F1 PRELIMINARY; PRT; 323 AA.
 AC Q9H1F1; PRELIMINARY; PRT; 323 AA.
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)

DE D644L1.1 (KREISLER (MOUSE) MAF-RELATED LEUCINE ZIPPER HOMOLOG).
GN KRL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramsey H.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035665; CAB75863.1; -
SQ SEQUENCE 323 AA; 35792 MW; A0F3C09F8936CB16 CRC64;

Query Match 22.7%; Score 85.5; DB 4; Length 323;
Best Local Similarity 23.1%; Pred. No. 6;
Matches 34; Conservative 9; Mismatches 19; Indels 85; Gaps 4;

OY 5 HHHHH-----GSMAS----- 15
DB 162 HHHHHQSPSSSAQOQLPTSHPGCPHATASATAGAGNGSVEDRFSDQLVMSV 221
OY 16 -----GDKKNVAOLKRRVRSIKDK-----AAELK 40
DB 222 RELNRHLRGFTKDEYIRLQKRRILKRGYAOQSCRYKRVQOKHLENEKTLIOQVEQL 281
OY 41 QEVSRLENEIEDLAKIGDLNNTSGIR 67
DB 282 QEVSRLENERDAVKYCKELAN-SGFR 307

RESULT 10
O31620 PRELIMINARY; PRT; 321 AA.
AC O31620:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE B-G (FRAGMENT).
GN B-G.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-91239571; PubMed-1903541;
RA Miller M.W., Goto R., Young S., Chirivella J., Hawke D., Miyada C.G.;
RT "Immunoglobulin variable-region-like domains of diverse sequence
within the major histocompatibility complex of the chicken."
Proc. Natl. Acad. Sci. U.S.A. 88:4377-4381(1991).
RL -1- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX
DOMAIN.
CC EMBL: M61862; AAA48627.1; -
DR InterPro: IPR003006; -
DR InterPro: IPR003600; -
DR Pfam: PF00047; Ig_1.
DR SMART: SM00410; Ig_like; 1.
KW MHC.
FT NON_TER 1
SQ SEQUENCE 321 AA; 36360 MW; 7F25BD29E39860C7 CRC64;

Query Match 22.5%; Score 85; DB 7; Length 321;
Best Local Similarity 26.8%; Pred. No. 6; 6;
Matches 22; Conservative 18; Mismatches 24; Indels 18; Gaps 2;

OY 7 HHHGSMASGDK-----NKVAOLKRRVRSIKRAELKQEVSRLENEIED 52
DB 185 HLAEKDLSADTKLAKLAVEQREAVERDSQKRYEKLGSATNLTQKLENEIEE 244
OY 53 LRAKIGDLNNTSGIRRAKLN 74

DB 245 VEKHKKI-----GIRAPRLKIH 262

RESULT 11
O9LAX2 PRELIMINARY; PRT; 479 AA.
AC O9LAX2:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PSPA (FRAGMENT).
GN PSPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BG7817;
RA Hollingshead S.K., Becker R., Briles D.E.;
RT "Diversity of PSPA: mosaic genes and evidence for past recombination
in Streptococcus pneumoniae."
Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF071826; AAF27719.1; -
DR InterPro: IPR000533; -
DR InterPro: IPR002965; -
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR01194; TROPOMYOSIN.
FT NON_TER 479
SQ SEQUENCE 479 AA; 53257 MW; B9C0D2CA15DE3654 CRC64;

Query Match 22.5%; Score 85; DB 2; Length 479;
Best Local Similarity 34.9%; Pred. No. 9; 9;
Matches 15; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

OY 17 DKNVAOLKRRVRSIKRAELKQEVSRLENEIEDLAKIGD 59
DB 215 ELQNRVADLEKEIDAETVADLEKEVAKLEKVEGRFESDSE 257

RESULT 12
O9LAX3 PRELIMINARY; PRT; 480 AA.
AC O9LAX3:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PSPA (FRAGMENT).
GN PSPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BG7561;
RA Hollingshead S.K., Becker R., Briles D.E.;
RT "Diversity of PSPA: mosaic genes and evidence for past recombination
in Streptococcus pneumoniae."
Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF071824; AAF27718.1; -
DR InterPro: IPR000533; -
DR InterPro: IPR002965; -
DR PRINTS: PR01217; PRICHEXTENS.
DR PRINTS: PR01194; TROPOMYOSIN.
FT NON_TER 480
SQ SEQUENCE 480 AA; 53043 MW; DA013C9E0190D7A0 CRC64;

Query Match 22.5%; Score 85; DB 2; Length 480;
Best Local Similarity 34.9%; Pred. No. 9; 9;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:13 ; Search time: 86.32 Seconds
(Without alignments)
239,490 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809
Sequence: 1 AEIYNKQKNKVDLYGKAVGL.....NKLGVGSDPTVAVGIYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

A_Geneseq_0601:*

- 1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	20	AAV42550
2	1792	99.1	340	17	AAW92998
3	1051	58.1	367	20	AAV42549
4	1051	58.1	367	20	AAV34058
5	1051	58.1	367	21	AAV57356
6	1029	56.9	377	20	AAV34057
7	1029	56.9	377	21	AAV57355
8	858.5	47.5	323	20	AAV34056
9	858.5	47.5	323	21	AAV57354
10	276	15.3	89	18	AAW2781
11	170.5	9.4	353	20	AAV26025

12	162.5	9.0	353	20	AAV26023
13	162	9.0	353	11	AAW06038
14	157	8.7	342	16	AAW07073
15	157	8.7	342	18	AAW21742
16	157	8.7	363	18	AAW07076
17	157	8.7	363	18	AAW21743
18	152.5	8.4	343	20	AAV26024
19	150.5	8.3	343	18	AAW21744
20	147.5	8.2	309	16	AAW07072
21	147.5	8.2	309	18	AAW21741
22	144.5	8.0	341	11	AAW07043
23	140	7.7	372	16	AAW06607
24	138.5	7.7	342	16	AAW66879
25	138.5	7.7	342	16	AAW66880
26	138.5	7.7	361	11	AAW05999
27	138.5	7.7	361	16	AAW66877
28	138	7.6	390	18	AAW21678
29	137.5	7.6	341	11	AAW07042
30	135.5	7.5	350	10	AAW90100
31	135.5	7.5	350	13	AAW27484
32	135.5	7.5	410	22	AAW19944
33	133	7.4	362	10	AAW90098
34	131.5	7.3	363	16	AAW66878
35	130	7.2	328	11	AAW6036
36	121	6.7	1122	16	AAW64927
37	119.5	6.6	326	11	AAW06037
38	118	6.5	730	21	AAW29582
39	118	6.5	752	21	AAW29581
40	116.5	6.4	185	19	AAW56320
41	114.5	6.3	707	19	AAW41523
42	113	6.2	467	21	AAW16023
43	111.5	6.2	373	11	AAW06035
44	111.5	6.2	934	17	AAW92289
45	111.5	6.2	972	17	AAW92288

ALIGNMENTS

RESULT	1	
AAV42550		
ID	AAV42550 standard; Protein; 362 AA.	
XX		
AC	AAV42550:	
XX		
DT	20-DEC-1999 (first entry)	
XX		
DE	E. coli wild-type ompF protein.	
KW	Bacteria; attenuation; deletion; mutant; vaccine; immune response;	
KW	Gram negative; infection; diarrhoea; food poisoning; typhoid;	
KW	salmonellosis; gonorrhoea; gastroenteritis; whooping cough.	
XX		
OS	Escherichia coli.	
XX		
PN	W09949026-A1.	
XX		
PD	30-SEP-1999.	
XX		
PF	25-MAR-1999; 99MO-GB00935.	
XX		
PR	25-MAR-1998; 98GB-0006449.	
XX		
PA	(PEPT-) PEPTIDE THERAPEUTICS LTD.	
XX		
PI	Chatfield SN;	
XX		
DR	WPI; 1999-580447/49.	
XX		
DR	N-PSDB; AA222887.	
XX		
PT	New attenuated bacteria useful as vaccines for protecting against	
XX	Infections -	

OMP protein of Pa
Class II outer mem
Mature Class 2 por
Neisseria meningit
Fusion Class 2 por
Neisseria meningit
OMP protein of Pa
Neisseria meningit
Meningococcal grou
Neisseria meningit
P3 gene product of
Gonococcal porin-5
H. influenzae prot
H. influenzae prot
P2 gene product of
H. influenzae P2 e
Haemophilus influe
P3 gene product of
Protein IB of N. g
Moraxella catarrha
P2 antigen of Haem
pNv-2 fusion const
Class IB outer mem
Cytohesin protein
Class IA outer mem
Arabidopsis thalia
Arabidopsis thalia
Haemophilus paraga
Blood group antige
E. coli proliferat
Class I outer memb
Cycloisomaltooligo
Cycloisomaltooligo

KW salmonellosis; gonorrhoea; gastroenteritis; whooping cough.
 XX Escherichia coll.
 XX MO9949026-A1.
 XX 30-SEP-1999.
 XX 25-MAR-1999; 99MO-GB00935.
 XX 25-MAR-1998; 98GB-0006449.
 XX (PEPT-) PEPTIDE THERAPEUTICS LTD.
 XX Chalfield SN;
 XX WPI; 1999-580447/49.
 XX N-PSDB; AA222885.
 XX New attenuated bacteria useful as vaccines for protecting against
 PT infections -
 XX PS Disclosure; Page 57-58; 69pp; English.
 XX This sequence represents the E. coli wild-type ompC protein. The coding
 CC sequence of the ompC gene was removed via PCR using primers TT7-TT10
 CC (AA222893-222896) to produce a non-reverting deletion mutation. The
 CC mutant ompC gene (AA222886) was used in the production of a bacterium
 CC attenuated by a non-reverting mutation in each of the ompC gene, the arcC
 CC gene (AA222883, AA222884), and the ompF gene (AA222887, AA222888). The
 CC mutant bacteria provide immunogenic activity with reduced virulence and
 CC thus can be used as a vaccine for raising an immune response against a
 CC variety of bacteria in a mammalian host. Such vaccines can provide
 CC protection against e.g., E. coli (a cause of diarrhoea in humans),
 CC Salmonella typhimurium (the cause of salmonellosis in several animal
 CC species), S. typhi (the cause of human typhoid), S. enteritidis (a cause
 CC of food poisoning in humans), S. choleraesuis (a cause of salmonellosis
 CC in pigs), S. dublin (a cause of both a systemic and diarrhoeal disease in
 CC cattle, especially of new-born calves), Haemophilus influenzae (a cause
 CC of meningitis), Neisseria gonorrhoeae (a cause of gonorrhoea), Yersinia
 CC enterocolitica (the cause of a spectrum of disease in humans ranging
 CC from gastroenteritis to fatal septicemic disease), Bordetella pertussis
 CC (the cause of whooping cough), and Brucella abortus (a cause of abortion
 CC and infertility in cattle and a condition known as undulant fever in
 CC humans).
 CC Sequence 367 AA;
 XX
 XX
 Query Match 58.1%; Score 1051; DB 20; Length 367;
 Best Local Similarity 58.9%; Pred. No. 9.6e-83;
 Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
 Oy 1 AEYNKDGNDVLDGKAVGLHYFSKNGENSYGNGDWTYRLGFKGTQINSDLTGCGQ 60
 Db 22 aeynknkgndldyqkvdglnhts-----dnkdvqgdqlymrlygkgetqldqlytgyqg 76
 Oy 61 WEYNFOGNNSEGADAOGTGNKTRLAFLKLYADVGSFDYGRNRYGVVYDALGTYDMLPERGG 120
 Db 77 weydlqgnsaen---emswtrvafaglkfdvgsfdygrnrygvvydvtswtdvlpefgy 133
 Oy 121 DTVASDFFVGRVGVATYRNSNFFGLVDGLNFAVOYLGNKNERDT-----A 166
 Db 134 dtygsdtdfmqgrngfatyrntdftglvdglnfavdyqgknpsqegftsgvtngrda 193
 Oy 167 RRSNGDGVGSISYEVGFGVGVAGVADRTNLOE-AQPLGNGKAKQOMAGLTYDANNI 225
 Db 194 lrtqgdvvgvgsilydyegfglgaissskrtcdagntaaylmgdraeyleyglkydanni 253
 Oy 226 YLAANYEETRNATPITKNTFTSGFANKTODVLLVAQYQDFGFRPSIAYKSKAKDV-E 284
 Db 254 ylaaeyqctynatrv-----gsigwankeqnfavvayqyfgfqlrpsalaylqskgknlgr 308

Oy 285 GIGDVLNVEFGATYENKMSYVYDIINOISDNKL-----GVSGDPTVAVGIYQF 340
 Db 309 gyddedllkyvdvagalyfknmstlydykxlnlld-dnqftrdaglnclnvalaylyyqt 367
 RESULT 4
 ID AAY34058 standard; protein; 367 AA.
 AC AAY34058;
 XX 23-NOV-1999 (first entry)
 DE E. coli outer membrane protein c precursor.
 XX
 XX Ulcerative colitis; histone; H1-like antigen; porin antigen;
 KW Bacteroides antigen; IBD; PANCA; inflammatory bowel disease;
 KW diagnosis; perinuclear anti-neutrophil cytoplasmic antibody;
 KW outer membrane protein c precursor.
 XX Escherichia coll.
 XX OS MO9945955-A1.
 XX PN 16-SEP-1999.
 XX 12-MAR-1999; 99MO-US05492.
 XX 12-MAR-1998; 98US-0041889.
 XX (REGC) UNIV CALIFORNTA.
 PA Braun J, Cohavy O;
 XX WPI; 1999-551215/46.
 XX Use of histone H1, porin or Bacteroides antigens as targets for the
 PT diagnosis, prevention and treatment of ulcerative colitis -
 XX Claim 8; Fig 11; 134pp; English.
 CC The invention provides a method for the diagnosis, prevention and
 CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a
 CC porin antigen or a Bacteroides antigen as a target antigen. The novel
 CC method of diagnosing UC in a subject suspected of having inflammatory
 CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
 CC (2) contacting the sample with a histone H1-like antigen, or perinuclear
 CC anti-neutrophil cytoplasmic antibody (pANCA)-reactive fragment, to form a
 CC complex of the histone H1-like antigen, or the pANCA-reactive fragment,
 CC and antibody to the histone H1-like antigen; and (3) detecting the
 CC presence or absence of the complex; where the presence of the complex
 CC indicates that the subject has UC. The pANCA-reactive histone H1-like
 CC antigen, porin antigen and Bacteroides antigen are useful in the
 CC diagnosis, prevention and treatment of UC. The methods can also be used
 CC for identifying agents useful for treating UC. The present sequence
 CC represents a E. coli outer membrane protein c precursor.
 CC Sequence 367 AA;
 XX
 XX
 Query Match 58.1%; Score 1051; DB 20; Length 367;
 Best Local Similarity 58.9%; Pred. No. 9.6e-83;
 Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
 Oy 1 AEYNKDGNDVLDGKAVGLHYFSKNGENSYGNGDWTYRLGFKGTQINSDLTGCGQ 60
 Db 22 aeynknkgndldyqkvdglnhts-----dnkdvqgdqlymrlygkgetqldqlytgyqg 76
 Oy 61 WEYNFOGNNSEGADAOGTGNKTRLAFLKLYADVGSFDYGRNRYGVVYDALGTYDMLPERGG 120
 Db 77 weydlqgnsaen---emswtrvafaglkfdvgsfdygrnrygvvydvtswtdvlpefgy 133
 Oy 121 DTVASDFFVGRVGVATYRNSNFFGLVDGLNFAVOYLGNKNERDT-----A 166


```
XX DR WPI; 1999-551215/46.
XX
XX PT Use of histone H1, porin or Bacteroides antigens as targets for the
XX diagnosis, prevention and treatment of ulcerative colitis
XX PS Claim 8; Fig 11; 134pp; English.
XX
XX The invention provides a method for the diagnosis, prevention and
XX treatment of ulcerative colitis (UC) using histone H1-like antigen, a
XX porin antigen or a Bacteroides antigen as a target antigen. The novel
XX method of diagnosing UC in a subject suspected of having inflammatory
XX bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
XX (2) contacting the sample with a histone H1-like antigen, or perinuclear
XX anti-neutrophil cytoplasmic antibody (PANCA)-reactive fragment, to form a
XX complex of the histone H1-like antigen, or the PANCA-reactive fragment,
XX and antibody to the histone H1-like antigen; and (3) detecting the
XX presence or absence of the complex; where the presence of the complex
XX indicates that the subject has UC. The PANCA-reactive histone H1-like
XX antigen, porin antigen and Bacteroides antigen are useful in the
XX diagnosis, prevention and treatment of UC. The methods can also be used
XX for identifying agents useful for treating UC. The present sequence
XX represents a E. coli outer membrane protein F precursor.
XX
XX Sequence 323 AA;
XX
XX Query Match 47.5%; Score 858.5; DB 20; Length 323;
XX Best Local Similarity 57.2%; Pred. No. 3.2e-66;
XX Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;
XX
XX QY 1 AEYKDGKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTVARLGFKEGTQINSDLTGYGQ 60
XX DB 22 aeaynkdkgnkldlygkvqgdlhyfsdnsek-----dgdqsyarlqfkgctqindqltygq 76
XX QY 61 WEYNQGNNSGADAGQGNKTRLAFLAGLKAVDVSFDYGRNYGVYDALGYTDMLEPFGG 120
XX DB 77 weynqgnnseskgqsw--rlaflaglkfadygsfdygrnygmvydmlpefgy 134
XX QY 121 DT-AVSDPEFYGRVGVATYRNSNFFGLVDGLNFAVQVLGKNE-----RDTRAR 167
XX DB 135 dytlnadnfmtrngvaytlyntdfjgvlnglnfavqyqgnnegasngqegtlngrd--vr 193
XX QY 168 RSNQDGVGSGSISYEX-EGFGIVGAYGAADRNLQLEAPLGNKKAQOMATGLKYDANNIY 226
XX DB 194 hengdgwglstctydlgmffsagaaytsdtrndqynhtaaggdkadawtqglkydanilly 253
XX QY 227 LAANGETRMAPTPTNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIAYTYSKAKDVEGCI 286
XX DB 254 latmysetrmtlptfgd---sdgyavanktqnfevtagyqfdfiglilpravsflnmskgydlhaa 310
XX
XX QY 287 GDVD 290
XX DB 311 gpad 314
XX
XX RESULT 9
XX ID AAY57354
XX AC AAY57354;
XX
XX DE 13-JUN-2000 (first entry)
XX
XX E. coli outer membrane protein F precursor.
XX
XX KW Ulcerative colitis; inflammatory bowel disease; porin antigen; MAAb;
XX PANCA; perinuclear anti-neutrophil cytoplasmic antibody;
XX histone H1; outer membrane protein F precursor.
XX
XX OS Escherichia coli.
XX
XX PN US6033864-A.
```

```
XX PD 07-MAR-2000.
XX PF 12-MAR-1998; 9805-0041889.
XX PR 12-APR-1996; 9605-0057846.
XX PR 11-APR-1997; 9705-0837058.
XX
XX (REBC ) UNIV CALIFORNIA.
XX
XX PI Cohavy O, Braun J;
XX
XX DR WPI; 2000-255695/22.
XX
XX PT Diagnosing ulcerative colitis or susceptibility, by detecting complex
XX formation between microbial porin antigen and perinuclear
XX anti-neutrophil cytoplasmic autoantibodies
XX
XX PS Claim 1; Fig 10; 49pp; English.
XX
XX The invention provides a method for diagnosing ulcerative colitis in a
XX subject suspected of having inflammatory bowel disease. The method
XX comprises reacting a patient sample with a porin antigen that is
XX immunologically reactive with PANCA (perinuclear anti-neutrophil
XX cytoplasmic antibodies) and detecting formation of a Ag-PANCA complex
XX as indicative of ulcerative colitis. The method is used to diagnose
XX ulcerative colitis or susceptibility to it. The present sequence
XX represents a E. coli outer membrane protein F precursor.
XX
XX Sequence 323 AA;
XX
XX Query Match 47.5%; Score 858.5; DB 21; Length 323;
XX Best Local Similarity 57.2%; Pred. No. 3.2e-66;
XX Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;
XX
XX QY 1 AEYKDGKNKVDLYGKAVGLHYFSKNGENSYGGNGDMTVARLGFKEGTQINSDLTGYGQ 60
XX DB 22 aeaynkdkgnkldlygkvqgdlhyfsdnsek-----dgdqsyarlqfkgctqindqltygq 76
XX QY 61 WEYNQGNNSGADAGQGNKTRLAFLAGLKAVDVSFDYGRNYGVYDALGYTDMLEPFGG 120
XX DB 77 weynqgnnseskgqsw--rlaflaglkfadygsfdygrnygmvydmlpefgy 134
XX QY 121 DT-AVSDPEFYGRVGVATYRNSNFFGLVDGLNFAVQVLGKNE-----RDTRAR 167
XX DB 135 dytlnadnfmtrngvaytlyntdfjgvlnglnfavqyqgnnegasngqegtlngrd--vr 193
XX QY 168 RSNQDGVGSGSISYEX-EGFGIVGAYGAADRNLQLEAPLGNKKAQOMATGLKYDANNIY 226
XX DB 194 hengdgwglstctydlgmffsagaaytsdtrndqynhtaaggdkadawtqglkydanilly 253
XX QY 227 LAANGETRMAPTPTNKFTNTSGFANKTQDVLVLAQYQFDGLRPSIAYTYSKAKDVEGCI 286
XX DB 254 latmysetrmtlptfgd---sdgyavanktqnfevtagyqfdfiglilpravsflnmskgydlhaa 310
XX
XX QY 287 GDVD 290
XX DB 311 gpad 314
XX
XX RESULT 10
XX ID AAM27781
XX AC AAM27781;
XX
XX DE 21-JUL-1998 (first entry)
XX
XX Amino acid sequence of an outer membrane protein F precursor.
XX
XX KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
XX Staphylococcal gene; regulatory element; bacterial gene expression;
```

Query Match	Best Local Similarity	Score	DB	Length
Matches 85; Conservative 41; Mismatches 132; Indels 109; Gaps 17;	9.4%;	170.5;	20;	353;
1 AEIYKDKNKKVDLYGKAVGLHYFSKNGENSXGNGDMTY--ARLGFKEQINSDLGY 58	23.2%;	7.7e-07;		
21 avyymegtkvelggrvsiiaeqscsnkkdqkqhgsirngsfrnkvtmhnldgqyal 80	41;	132;	109;	Gaps 17;
59 GOMEYNFOGNNSEGADAQTGN-----KTRLAFAGLKYADVGSFDYG----- 99				
81 gyycerflinkldgqneklsgfsgitckllyagqgnelgsatqlqtkladdkstaed 140				
100 RNYGVYVALGTTDLPEFGGDTAVSDPEFVGRVGVCAVTRNSNFFGLVDGILNFAVQILG 159				
141 keygylek-----nsyiplegnalay-----tykg-----legltlqasyv- 176				
160 KNERPTARNSGDDGVGSISYSEGEFGIVGAVGADPRNLGEAQFLGNGKRAEQMATGK 219				
177 -----fggnfnfsdy-----itcqkvsnavqvyak 201				
220 YDANNIYLAANGETRNATPTITNKFTNTSGFANKTQDY---LLVAQYOF-DFGLRPSI- 273				
202 ydamniivgfygr-----tnykaqgaktqyngalactgynhfdldgllistsid 250				
274 AYTRSKAKVDGEGIDVDLVNPF-EVGATYTFPNKKNSTY--VDY-IINQIDSDNKLGVGSD 329				
251 gyaktknk-----adkhekryfvsppgfygelmtdnlgnlkyerlnsvdgsgekv---re 302				
330 DTVAVGI 336				
303 havlfgl 309				

XX	AAV26023:
AC	
XX	29-SEP-1999 (first entry)
DT	
XX	OmpH protein of Pasteurella multocida X-73.
DE	
XX	Outer membrane protein H; ompH gene; fowl cholera; molecular sieve;
KW	Pasteurella multocida X-73; porin; avian; vaccine; veterinary medicine;
KW	bacterial porin; H. influenzae porin P2.
XX	
OS	Pasteurella multocida.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..20
FT	/label= Signal_peptide
FT	/note= "Hydrophobic amino acids"
FT	Protein 21..353
FT	/note= "mature OmpH protein"
FT	Cleavage-site 7..9
FT	/note= "Present within the signal peptide"
FT	Cleavage-site 9..11
FT	/note= "Present within the signal peptide"
FT	Cleavage-site 12..14
FT	/note= "Present within the signal peptide"
FT	Cleavage-site 18..20
FT	/note= "Present within the signal peptide"
FT	Peptide 81..110
FT	/label= MAP-I2.peptide
FT	/note= "Multiple Antigen Peptide derived from loop 2"
FT	Peptide 211..238
FT	/label= MAP-I5.peptide
FT	/note= "Multiple Antigen Peptide derived from loop 5"
PN	WO929724-A2.
XX	
PD	17-JUN-1999.
XX	
PE	08-DEC-1998; 98MO-US25990.
PR	21-JUL-1998; 98US-0120051.
PR	08-DEC-1997; 97US-0067957.
XX	
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.
XX	
P1	Glisson JR, Luo Y;
DR	MP1: 1999-385575/32.
DR	N-PSDB: AAX80609.
XX	
PT	DNA encoding Pasteurella multocida outer membrane protein H, useful
PS	for delivering peptide epitopes for vaccines against fowl cholera
XX	
PS	Claim 10; Fig 4; 11pp; English.
XX	
CC	The present amino acid sequence is that of the avian Pasteurella
CC	multocida X-73, outer membrane protein H (OmpH). This protein is a
CC	porin, which is a major outer membrane protein that is involved in
CC	pore formation. They serve as molecular sieves that allow polar solutes
CC	to pass through, but excludes non-polar molecules of comparable sizes.
CC	The amino acid composition is typical of non-specific bacterial porins
CC	with highly negative hydropathy index, high glycine content, low proline
CC	content and lack of cysteine. This amino acid sequence shows homology of
CC	about 38% to bacterial porins, like that of H. influenzae porin P2.
CC	Multiple antigen peptides are synthesized from the loop 2 and loop 5
CC	regions of the protein that can be used for immunisation of the fowls
CC	to induce heterologous protection against P.multocida infection. This
CC	sequence can be used for diagnostic assays for use in veterinary
CC	medicine, especially for the diagnosis of fowl cholera, caused by
CC	P.multocida. Peptides derived from the OmpH protein are useful in the
CC	production of protective vaccines for use in vertebrates, in particular
CC	for prevention of fowl cholera.
XX	

Seq	Sequence	353 AA;
Query Match	9.08; Score 162.5; DB 20; Length 353;	
Best Local Similarity	22.94; Pted. No. 3.8e-06;	
Matches	83; Conservative 39; Mismatches 150; Indels 91; Gaps	
Oy	1 AEINRDKQKVLDYKKAIVGLHYFSKNGKNGSNSYGGNGDMFYARLGFKGETQINSDLNGYQ 60	
Db	21 atyngdgctkxdvng-slrllkkkehergdlvng-----srysfkashdlgeglasalay 75	
Oy	61 WEYNFGNNGSEGADATQGNKTR-----LAFAGLKADYGSFDYGRNCGVT 106	
Db	76 telfsknvpyqvkddqgevvreyeveklgnvhwkrllyagfayeglgltlfgnqltl-- 133	
Oy	107 DALGVTDLPEFGGDDTAYSDDFFGRRGVATYNSNFPFGJNGDLNFAVGYLGKNERDFA 166	
Db	134 -----gddvglsd-----ykyfnsglnml----- 153	
Oy	167 RRSNGDVGSGISYEYEGFIVGAY---GAADRTNLOEAQ-----PLNGKKAQEWATGL 218	
Db	154 -ssgekalnfkasefngftlgagvysaadkqalrdrgfvgvaylnyknmgd---vgf 208	
Oy	219 KYDANNITLANY--GETRNPATPTNKFNTSGFRANKTQDVLVAQYQFDFGLRPSIAT 276	
Db	209 afeag---ysqkyvkqevqgnppaaqkv---FkdekakaImygaelsyaglaIgydva 260	
Oy	277 KSKAKDVEGIDVDLVNVEFGATYFFNKNMSTYVDYIIINOIDSNDKLGVSDDTVAVGI 336	
Db	261 qskvtinvdg-----kkralevglnydlndrakvytdfiwek--egpkgdvtlnrtvavgf 313	
Oy	337 VYQ 339	
Db	314 gylk 316	
RESULT 13		
AA06038		
ID	AA06038 standard; protein; 353 AA.	
XX		
AC	AA06038;	
XX		
CT	20-NOV-1990 (first entry)	
DE	Class II outer membrane protein of N.meningitidis.	
XX		
KW	Meningococcal disease; meningitis; vaccine;	
KW	Class I outer membrane protein; ds.	
XX		
OS	Neisseria meningitidis.	
XX		
PN	W09006696-A.	
XX		
PD	28-JUN-1990.	
XX		
PF	19-DEC-1989; 89WC-US05678.	
XX		
PR	26-JUN-1989; 89NL-0001612.	
XX		
PR	19-DEC-1988; 88NL-0003111.	
XX		
PR	06-JAN-1989; 89NL-0000030.	
XX		
PR	06-JAN-1989; 89NL-0000036.	
XX		
PA	(PRAK-) PRAXIS BIOLOGICS IN.	
XX	(VOLK-) RIJIKINSINST VOLKSGEZONDH.	
PI	SEID RC, PARADISO PR, POOLMAN JT, HOOGERHOUT P, WIERTZ EJ;	
XX		
PI	VANDERLEY P, HECKELS JE, CLARKE IN;	
XX		
DR	WPI; 1990-224326/29.	
XX		
PT	Meningococcus class I outer-membrane protein vaccine - useful to	
XX	immunise against meningococcal disease.	
XX		

PS Disclosure; : 120pp; English.

XX Peptides derived from outer membrane, encoding homo/heterologous
 CC product such as an antigen-flagelin fusion protein, are useful in
 CC eliciting a strong and wide ranging immune response against most
 CC serotypes.

XX Sequence 353 AA:

Query Match 9.0%; Score 162; DB 11; Length 353;
 Best Local Similarity 26.6%; Pred. No. 4.2e-06;
 Matches 79; Conservative 42; Mismatches 126; Indels 50; Gaps 16;

QY 41 ARIGFKEIOINSDLTGGMETNFQGNSEGADQTKNTRLAFLAKYADGSPDYGR 100
 DB 54 sklgfkgedelgnglkawle---qksiafgtnsgvgn---rqsflgk---ggfgtivr 104

QY 101 --NCGVYVDALGYTDMPEFGDTRVSDDFVCGVGVGATVTRNSNFRG--LVQGLNFAVQ 156
 DB 105 aglnlntvlykagadnvnawesgnt--edvlglgfgrvesrelsvrldpvrfgfsgsvq 162

QY 157 YLKG---NERDTRARNSGDCVGSISYEYEGFIVGAYGA-----ADR--TNLQ 200
 DB 163 yperdnandvdkykhktsaresyhaaglkynagffgqyagsfakyladlnldaeravanta 222

QY 201 EAQPLNGKKAEGWANGTKYDANNIY--LAANYGETRNATPTNKFTNTSGFANKTODVL 258
 DB 223 nshpv---kdyqvlrvvagyadandlysvagyaakn-----nevsgtkykhkhtq-va 273

QY 259 LVNQYQDFGLRPSIATKSKAKDVEGIDVDLVNFE---VCAITYYFNKNMSTYV 311
 DB 274 ataaytrfg-nvcpvryahgfkakvngkd---anyqddqvivgadydfskrtsalv 326

RESULT 14
 AAR70763
 ID AAR70763 standard; Protein: 342 AA.

XX AAR70763;

XX 24-AUG-1995 (first entry)

XX Mature class 2 porin expressed from gene cloned into pET-17b.

XX Group B porin protein; meningococcal; outer membrane protein;
 KW vaccine; meningitis; Neisseria meningitidis; PCR primer.

XX Synthetic.

XX MO9503413-A.

XX 02-FEB-1995.

XX 22-JUL-1994; 94MO-US08327.

XX 23-JUL-1993; 93US-0096182.

XX (NAVA-) NORTH AMERICAN VACCINE INC.
 PA (UYRO) UNIV ROCHEFELLER.

XX Blake MS, Hronowski LJ, Liang S, Pullen JK, Qi HL;
 PI Tal JY;

XX WPI: 1995-075239/10.
 DR N-PSDB; AAQ85391.

XX High expression of outer membrane meningococcal group B porin
 PT proteins - and fusion proteins in Escherichia coli, and
 PT purification method; for use in vaccines against Neisseria
 PT meningitidis and in research.

XX Example; Figure 9; 81pp; English.

XX The example concerns cloning of Class 2 porin from Group B
 CC Neisseria meningitidis strain BNCV M986 serotype 2a (see
 CC AA085300, AA085301). The plasmid pET-17b was used to express the
 CC class 2 porin. A plasmid was designed to yield a mature class 2
 CC porin. The mature class 2 porin was constructed by amplifying
 CC the pUC19-class 2 porin construct using the oligos AA085302 and
 CC AA085303. This strategy allowed the cloning of the amplified class 2
 CC porin into the NdeI and XhoI sites of the plasmid pET-17b thus
 CC producing mature class 2 porin. Std. PCR was conducted using the
 CC pUC19-class as the template and the two oligos. The PCR reaction
 CC yielded a 1.1kb product. The DNA obtd. was purified and digested
 CC with NdeI and XhoI and ligated to pET-17b, and used to transform
 CC E.coli. DNA from transformed clones was sequenced. AA085291/R70763
 CC show the nt and translated AA sequence of the mature class II
 CC porin gene cloned into the expression plasmid pET-17b.

XX Sequence 342 AA:

Query Match 8.7%; Score 157; DB 16; Length 342;
 Best Local Similarity 24.7%; Pred. No. 1.1e-05;
 Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLGKAVGLHFFKNGENSYGNGDMTY-----ARLGRGKQINSDLTGIGQW 61
 DB 3 vlygltkagvevsvrxdagtykagqgsktatqadfgsklgtkgqdelgngmkalwq1 62

QY 62 EYNQGNSEGADQTKNTRLAFLAKYADVGSFDYGR--NCGVYVDALGYTDMLEFG 119
 DB 63 e---qksiafgtnsgvgn---rqsflgk---ggfgrvrasnlnltvlyksgdnvnawesg 113

QY 120 GDTAVSDDFEYGVGVGATVTRNSNFRG--LVQGLNFAVQYLGK---NERDTRARNSGDCV 174
 DB 114 snt--edvlgfgrvesrelsvrldpvrfgfsgsvqyprdnandvdkykhktsare 171

QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNQEAQPLNGKKAQWMTGLKYD 221
 DB 172 syhaglkynagffgqyagsfakyladlnldaeravantaahpv---kdyqvlrvvagyad 228

QY 222 ANNIY--LAANYGETRNATPTNKFTNTSGFANKTODVLVAQYQDFGLRPSIATYKSK 279
 DB 229 andlysvagyaakn-----nevsgtkykhkhtqyaaataayrfg-nvcpvryahg 282

QY 280 AKDVEGIDVDLVNFE---FEVGATYFFNKNMSTYV 311
 DB 283 kakvngvkd---anyqddqvivgadydfskrtsalv 315

RESULT 15
 AAW21742
 ID AAW21742 standard; Protein: 342 AA.

XX AAW21742;

XX 09-MAR-1998 (first entry)

XX Neisseria meningitidis class 3 mature porin protein.

XX Porin protein; porB gene; outer membrane protein; MB3;
 KW meningitis; vaccine.

XX Neisseria meningitidis serogroup B strain 8765 (B:15:PI.3).

XX MO9728273-A1.

XX 07-AUG-1997.

XX 31-JAN-1997; 97WO-US01687.

XX 13-JUN-1996; 96US-0020440.

XX 01-FEB-1996; 96US-0010972.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:45 ; Search time 48.42 Seconds

(Without alignments)
141.869 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1 AEIYNKDGKNKVDLYGKAVGL.....NKLGVGSDDTVAVGIVQFA 341

Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1792	99.1	340	2	US-08-355-844-1
2	1792	99.1	340	5	PCT-US95-16126-1
3	1051	58.1	367	3	US-09-041-889-30
4	1029	56.9	377	3	US-09-041-889-29
5	858.5	47.5	323	3	US-09-041-889-28
6	157	8.7	342	1	US-08-096-182A-4
7	157	8.7	342	1	US-08-096-182A-4
8	157	8.7	342	3	US-08-798-760-4
9	157	8.7	342	5	PCT-US94-08327-4
10	157	8.7	363	1	US-08-096-182A-6
11	157	8.7	363	1	US-08-877-109-6
12	157	8.7	363	3	US-08-798-760-6
13	157	8.7	363	5	PCT-US94-08327-6
14	147.5	8.2	309	1	US-08-096-182A-2
15	147.5	8.2	309	1	US-08-877-109-2
16	147.5	8.2	309	3	US-08-798-760-2
17	147.5	8.2	309	5	PCT-US94-08327-2
18	138.5	7.7	342	4	US-08-096-181A-12
19	138.5	7.7	342	4	US-08-096-181A-14
20	138.5	7.7	342	5	PCT-US94-08326-12
21	138.5	7.7	342	5	PCT-US94-08326-14
22	138.5	7.7	361	4	US-08-096-181A-8
23	138.5	7.7	361	4	US-08-096-181A-10
24	138.5	7.7	363	5	PCT-US94-08326-10
25	108.5	6.0	511	1	US-08-480-604A-20
26	108.5	6.0	511	2	US-08-405-496A-20
27	108.5	6.0	511	2	US-08-405-496A-20

28	108.5	6.0	608	1	US-08-480-604A-21	Sequence 21, Appl
29	108.5	6.0	608	2	US-08-405-496A-21	Sequence 21, Appl
30	108.5	6.0	609	1	US-08-480-604A-30	Sequence 30, Appl
31	108.5	6.0	2366	1	US-08-480-604A-10	Sequence 10, Appl
32	108.5	6.0	2366	2	US-08-405-496A-10	Sequence 10, Appl
33	105	5.8	455	2	US-08-472-172-4	Sequence 4, Appl
34	103.5	5.7	1794	6	5183745-6	Patient No. 5183745
35	102.5	5.7	1489	6	5183745-2	Patient No. 5183745
36	100.5	5.6	1577	2	US-08-793-824-2	Sequence 2, Appl
37	100	5.5	1115	3	US-08-323-477-2	Sequence 2, Appl
38	99.5	5.5	432	2	US-08-472-172-6	Sequence 6, Appl
39	99.5	5.5	700	2	US-07-862-588B-2	Sequence 2, Appl
40	99	5.5	459	2	US-08-472-172-2	Sequence 2, Appl
41	98.5	5.4	617	1	US-08-361-920-29	Sequence 29, Appl
42	98.5	5.4	617	1	US-08-479-939-29	Sequence 29, Appl
43	98.5	5.4	617	1	US-08-483-432-29	Sequence 29, Appl
44	94.5	5.2	607	4	US-08-537-361E-10	Sequence 10, Appl
45	94.5	5.2	607	4	US-08-537-361E-10	Sequence 10, Appl
					US-08-537-361E-10	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-355-844-1
Sequence 1, Application US/08355844
Patent No. 5940307
GENERAL INFORMATION:
APPLICANT: Fischbary, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/355,844
FILING DATE: 14-DEC-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y.S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A29927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OMPF porin protein
US-08-355-844-1

Query Match 99.1%; Score 1792; DB 2; Length 340;
Best Local Similarity 99.4%; Pred. No. 2e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYKDKNKKVDLYGKAVGLHYFSKNGENSYGCGMDTYARLGKGETQINSDLTGQG 60
DB 1 AEIYKDKNKKVDLYGKAVGLHYFSKNGENSYGCGMDTYARLGKGETQINSDLTGQG 60
QY WEYNQGNNSGADQOTGNKTRLAFAGLKYADVGSFDYGRNMGVYDALGYTDMLEFGG 120
DB 61 WEYNQGNNSGADQOTGNKTRLAFAGLKYADVGSFDYGRNMGVYDALGYTDMLEFGG 120
QY 121 DTAYSDDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKNERDTPARRSNGDVGGSISY 180
DB 121 DTAYSDDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKNERDTPARRSNGDVGGSISY 180
QY 181 EYEGFGIYGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNITYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNITYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYTKSAKADVEGIGVDLVNFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYTKSAKADVEGIGVDLVNFEVGAT 300
QY 301 YFENKMSYVDYIIINOIDS DNKLGVGSDDTVAAGIYVQF 340
DB 301 YFENKMSYVDYIIINOIDS DNKLGVGSDDTVAAGIYVQF 340

RESULT 2
PCT-US95-16126-1
Sequence 1, Application PC/TUS9516126
GENERAL INFORMATION:
APPLICANT: Fischbarg, Jorge
APPLICANT: Czegledy, Ferenc
APPLICANT: Iserovich, Pavel
APPLICANT: Li, Jun
APPLICANT: Cheung, Min
TITLE OF INVENTION: A METHOD FOR PREDICTING PROTEIN
TITLE OF INVENTION: STRUCTURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10112-0228
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,844
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Tang, Henry Y. S.
REGISTRATION NUMBER: 29,705
REFERENCE/DOCKET NUMBER: A29927-50/29910
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-408-2586
TELEFAX: 212-765-2519
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 340 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..340
OTHER INFORMATION: OmpF porin protein
PCT-US95-16126-1

Query Match 99.1%; Score 1792; DB 5; Length 340;
Best Local Similarity 99.4%; Pred. No. 2e-159;
Matches 338; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AEIYKDKNKKVDLYGKAVGLHYFSKNGENSYGCGMDTYARLGKGETQINSDLTGQG 60
DB 1 AEIYKDKNKKVDLYGKAVGLHYFSKNGENSYGCGMDTYARLGKGETQINSDLTGQG 60
QY 61 WEYNQGNNSGADQOTGNKTRLAFAGLKYADVGSFDYGRNMGVYDALGYTDMLEFGG 120
DB 61 WEYNQGNNSGADQOTGNKTRLAFAGLKYADVGSFDYGRNMGVYDALGYTDMLEFGG 120
QY 121 DTAYSDDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKNERDTPARRSNGDVGGSISY 180
DB 121 DTAYSDDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKNERDTPARRSNGDVGGSISY 180
QY 181 EYEGFGIYGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNITYLAANYGETRNATPI 240
DB 181 EYEGFGIYGAGADRTNLOEAOPLGNGKKAQOMATGLKYDANNITYLAANYGETRNATPI 240
QY 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYTKSAKADVEGIGVDLVNFEVGAT 300
DB 241 TNKFTNTSGFANKTQDVLLVAQYOFDFGLRPSIATYTKSAKADVEGIGVDLVNFEVGAT 300
QY 301 YFENKMSYVDYIIINOIDS DNKLGVGSDDTVAAGIYVQF 340
DB 301 YFENKMSYVDYIIINOIDS DNKLGVGSDDTVAAGIYVQF 340

RESULT 3
US-09-041-889-30
Sequence 30, Application US/09041889
Patent No. 6033864
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Cohavy, Ofer
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-09-041-889-30

Query Match 58.1%; Score 1051; DB 3; Length 367;
Best Local Similarity 58.9%; Pred. No. 2.9e-90;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;

QY 1 AEYINKDGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 60
DB 22 AEYINKDGNKLDLYGKVDGLHYFS-----DNKVDGDQTYMRLGFGEQVTDQLTGYGQ 76
QY 61 WEYNFGNNSGADAOCTGNKTRLAFLAGLKYADYSPDYGRNRYGVYDALGYTDLPEFGG 120
DB 77 WEYQIGNSAEN---ENNSMTRVAFAGLKFQDYGSPDYGRNRYGVYDVTWTDVLPPEFGG 133
QY 121 DTAASDFEYGRVGVYATYRNSNPFGLVDGLNFAVOYLKNEBDT-----A 166
DB 134 DTYSNFMQORNGCATYNTDFGLVDGLNFAVOYQKGNPNBSGFTSGVTNNGRDA 193
QY 167 RRSNGGVGSISYSEYEGEIGVAYGADRTNIOE-AQPLNGKKAQMATGLKYDANNI 225
DB 194 LRONGGVGSISYDEGEIGVAYGADRTNIOE-AQPLNGKKAQMATGLKYDANNI 253
QY 226 YLANGETRNATPTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADY-E 284
DB 254 YLANQYTOTYRNATRV-----GSLGMANKAQNFVAVAOYQDFGLRPSLATLQSKKILGR 308
QY 285 GIDVDLVNFEVGYATYFNKNMSTYVDYIINOISDNKL-----GVGSDPTVAVGIVYQF 340
DB 309 GYDDEDILKYVDGATYFNKNMSTYVDYKINLDD-DNCFTRDAGINTDNIVALGLVYQF 367

RESULT 4
US-09-041-889-29
Sequence 29, Application US/09041889
Patent No. 6033864

GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes thereof, using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 377 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-29

Query Match 56.9%; Score 1029; DB 3; Length 377;
Best Local Similarity 56.9%; Pred. No. 3.4e-88;
Matches 209; Conservative 41; Mismatches 79; Indels 38; Gaps 9;

QY 1 AEYINKDGNKVDLYGKAVGLHYFSKNGENSYGNGDMTYARLGFGEQINSDLTGYGQ 60
DB 22 AEYINKDGNKLDLYGKVDGLHYFS-----DGDQSYARLGFGEQVTDQLTGYGQ 76
QY 61 WEYNFGNNSGADAOCTGNKTRLAFLAGLKYADYSPDYGRNRYGVYDALGYTDLPEFGG 120
DB 77 WEYQIGNSAEN---ENNSMTRVAFAGLKFQDYGSPDYGRNRYGVYDVTWTDVLPPEFGG 134
QY 121 DTAASDFEYGRVGVYATYRNSNPFGLVDGLNFAVOYLKNE-----RDTAR 167
DB 134 DTYSNFMQORNGCATYNTDFGLVDGLNFAVOYQKGNPNBSGFTSGVTNNGRDA 193
QY 168 RRSNGGVGSISYSEYEGEIGVAYGADRTNIOE-AQPLNGKKAQMATGLKYDANNI 226
DB 194 LRONGGVGSISYDEGEIGVAYGADRTNIOE-AQPLNGKKAQMATGLKYDANNI 253
QY 227 YLANGETRNATPTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSKADY-E 284
DB 254 YLANQYTOTYRNATRV-----GSLGMANKAQNFVAVAOYQDFGLRPSLATLQSKKILGR 308
QY 285 GIDVDLVNFEVGYATYFNKNMSTYVDYIINOISDNKL-----GVGSDPTVAVGIVYQF 340
DB 311 GADNPAGVDKDLKYKADIGATYFNKNMSTYVDYKINLDD-DNCFTRDAGINTDNIVALGLVYQF 370

RESULT 5
US-09-041-889-28
Sequence 28, Application US/09041889
Patent No. 6033864

GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes thereof, using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997

APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-28

Query Match 47.5%: Score 858.5; DB 3; Length 323;
Best Local Similarity 57.2%: Pred. No. 2.1e-72;
Matches 174; Conservative 33; Mismatches 72; Indels 25; Gaps 7;

QY 1 AEVYKDKNKVLYGKAGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60
DB 22 AEVYKDKNKDLVYKVGVLHGFSDNSAK-----DGDQSYARLFGKGTQINDLTGYGQ 76
QY 61 WEYNQGNNSCADQOTGKTRFLAFAGLKYADVGSFYDGRNRYVYDALGYTDLMPERG 120
DB 77 WEYNQANNTESSKQSW--TRLAFAGLKFDYGSFYDGRNRYVYDALGYTDLMPERG 134
QY 121 DT-AVSDDFVGRGVATYRNSNFFGLVDGLNFAVQYLKNE-----RDTAR 167
DB 135 DSYTANDNFMTRANGVATRYNTDFEGLVNGLNFAVQYOGNNEGASNGEGTNNGRD-VR 193
QY 168 RNSGDGVGSSISEY-EGFGIVGAYGAADRINLQENQPLNGSKKAEQWATGLKYANNY 226
DB 194 HENGDOMGLSTYYDLGMGFSAGAAVTSRDNDQVNHFAAGDGDADWATGAKKYANNY 253
QY 227 LAANGERNATPTINKFTNTSGFANKQDVLLVQYOPDFGLRSIATRSKADVECI 286
DB 254 LRTMSETNNMTPFGD---SDYAVANKTQNEFVTAQYDFGLRPVSPLSKGRDLAA 310
QY 287 GPVD 290
DB 311 GGAD 314

RESULT 6
US-08-096-182A-4
Sequence 4, Application US/08096182A
Patent No. 5439808
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Ol, Huihan L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hironowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porlin Proteins from Neisseria Meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,182A
FILING DATE: 23-JUL-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Esmond, Robert W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060000
TELEPHONE: (202) 371-2500
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-182A-4

Query Match 8.7%: Score 157; DB 1; Length 342;
Best Local Similarity 24.7%: Pred. No. 6.1e-07;
Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VLYGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKGTQINSDLTGYGQ 61
DB 3 VLYGTITAGVEYSKXAGTYKAGGSKSKATQIADFGSKIGFGQDGLGMAIQL 62
QY 62 EYFQGNNSCADQOTGKTRFLAFAGLKYADVGSFYDGR--NYGVYDALGYTDLMPERG 119
DB 63 E--OKASIGATNSGMGN--RQSFGLK---GGFTVRAGLNTVLKDSGQNVAMSEG 113
QY 120 GDTAVSDDFVGRGVATYRNSNFFG--LYDGLNFAVQYLK---NBRDTARRSNGDV 174
DB 114 SMT--EDVLGLTTRVESREISVXKDSVPVFGSGSYQYVRDANVDKHKHRSSE 171
QY 175 GGSISEYEFGFIVGAYGA-----ADR--TNLQENQPLNGSKKAEQWATGLKYD 221
DB 172 SYHAGLKEENAGFFQYVAGSFAKYADLWTDABRVAVNTANAHV---KDYQVHRVAGD 228
QY 222 ANNTY--LAANGERNATPTINKFTNTSGFANKQDVLLVQYOPDFGLRSIATRSK 279
DB 229 ANDLYVYVAGQYEAKN---NEVGSTKGRKHEQTQVAATAAIFRG-NVTPRVSIANGF 282
QY 280 AKDVEGIGVDLVNY---FEVGATYFENKNSFTYV 311
DB 283 KAKVNGVD---ANTQYQYIVGADYDFSKRTSALV 315

RESULT 7
US-08-877-109-4
Sequence 4, Application US/08877109
Patent No. 5747287
GENERAL INFORMATION:
APPLICANT: Blake, Milan S.
APPLICANT: Tai, Joseph Y.
APPLICANT: Ol, Huihan L.
APPLICANT: Liang, Shu-Mei
APPLICANT: Hironowski, Lucjan J.J.
APPLICANT: Pullen, Jeffrey K.
TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B
TITLE OF INVENTION: Porlin Proteins from Neisseria meningitidis
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Ave., NW
CITY: Washington


```

1  APPLICANT: 12103 Indian Creek Court
2  APPLICANT: Beltsville, Maryland 20705
3  APPLICANT: United States of America
4  APPLICANT: Tal, Joseph Y.
5  APPLICANT: Qi, Huijin L.
6  APPLICANT: Liang, Shu-Mei
7  APPLICANT: Hironowski, Lucjan J. J.
8  APPLICANT: Pullen, Jeffrey K.
9  TITLE OF INVENTION: Method for the High Level
10 TITLE OF INVENTION: Expression,
11 TITLE OF INVENTION: Purification and Refolding of the Outer Membrane
12 TITLE OF INVENTION: Group B
13 TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis
14 NUMBER OF SEQUENCES: 23
15 CORRESPONDENCE ADDRESS:
16 ADDRESSEE: Sterne, Kessler, Goldstein & Fox
17 STREET: 1100 New York Ave., Suite 600
18 CITY: Washington
19 STATE: D.C.
20 COUNTRY: USA
21 ZIP: 20005-3934
22 COMPUTER READABLE FORM:
23 MEDIUM TYPE: Floppy disk
24 COMPUTER: IBM PC compatible
25 OPERATING SYSTEM: PC-DOS/MS-DOS
26 SOFTWARE: Patentln Release #1.0, Version #1.25
27 CURRENT APPLICATION DATA:
28 APPLICATION NUMBER: PCT/US94/08327
29 FILING DATE: Herewith
30 CLASSIFICATION:
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER: US 08/096,182
33 FILING DATE: 23 July 1993
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Esmond, Robert W.
36 REGISTRATION NUMBER: 32,893
37 REFERENCE/DOCKET NUMBER: 1438, 006PC00
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: (202) 371-2600
40 TELEFAX: (202) 371-2540
41 INFORMATION FOR SEQ ID NO: 4:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 342 amino acids
44 TYPE: amino acid
45 TOPOLOGY: linear
46 MOLECULE TYPE: protein
47 PCT-US94-08327-4

```

[illegible]

```
QY      280 AKDVEGIGDVLVNY----FEVGATYFNKNMSTYV 311
          | : | | | | | | |
Db      283 KAKVNGVKD---ANYQYDQYIVGADYDESKRTSALV 315
```

```

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
104
```

	Query Match	8.7%	Score 157;	DB 1;	Length 363;	
	Best Local Similarity	24.7%	Pred. No. 6.7e-07;			
	Matches	83;	Conservative	45;	Mismatches 150;	Indels 58; Gaps 16;
Oy	11	VDLYGKAVGLHYEFSKGENBNTYSGNDMTY-----ARLGFGEQTOINSDLTGYGQM	61			
Db	24	VLYLTITAGVEVSVHKDAGTIGKAGGKSKTATYADFGSKIGKGFQGDLDGKMAIWL	83			
Oy	62	EYNFGNNSEGADQGTGKTKRLAFGLKYRADGSPDYGR--NYGVVYDALGYTDMLPREG	119			
Db	84	E---QKASITAGINSWGN--RQSFITGLK----GGGCTYRAGLNTVYLDSDGNVNAMEG	134			
Oy	120	GDTAVSDPEFVGRCVGVATYRNSNFFG--LYDGLNFAVOYLK---NEEDTARSRNGDGV	174			
Db	135	SNT---EDVLGLGTTIGRVSRREISVYRDSPEVFGSGSVQVYVERDANADVKKKHKTRSSRE	192			
Oy	175	GGSISEYEGFVIGAYGA-----ADR--TNLOEAPLNGGKKAQOMATGLTKLD	221			
Db	193	SYHAKLKTENGFQOYAGSPFAKYADLMTDRAVAVNTANMHPV---KDYQHNRYVAGCD	249			
Oy	222	ANNIY--LAANYGETRNPATPTNKFTNTSGFANKTQDVLVAOYDFDGLRPSIATYTKK	279			

Db 250 ANDLYSVAGQYERAKN-----NEVGSTGKKEHQVOVATAATAYRGG-NWTPRVSYAHGF 303

QY 280 AKDVEGIGVDLVN-----FEVATYFFNKNMSTYV 311

Db 304 KAKVNGVKD---ANYQYDOYIVGADYDFSKRTSALV 336

RESULT 11

US-08-877-109-6
; Sequence 6, Application US/08877109

; Patent No. 5747287

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tal, Joseph Y.

; APPLICANT: Qi, Hulin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/877,109

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/431,264

; FILING DATE: 28-APR-1995

; APPLICATION NUMBER: 08/096,182

; FILING DATE: 23-JUL-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-877-109-6

Query Match 8.7%; Score 157; DB 1; Length 363;

Best Local Similarity 24.7%; Pred. No. 6,7e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLYGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKGTQINSDLTGQW 61

Db 24 VTLYGKIKAGVEVSRKVDAGTYKRAQGGKSKTATQIADFGSKIGFGQEDLGNMKAIMWL 83

QY 62 EYNFGNNSGADAGQGNKRLAFAGLKYADVGSFDYGR--NNGVYVDALGYDMLPEFG 119

Db 84 E--QKASTAGTNSGMCN--ROSFILK---GGFGTVRAGNLTNVLKDSGDVNWAMESG 134

QY 120 GDTAVSDDFVGVGVAVIYRNSNFG--LVLDLNFVAVYGLK---NERDTARNSGDGV 174

Db 135 SNT--EDVLGLGTIGRVESREISVRDPSVFAGFSGSVQVPRDNANDVDYKHTKSSRE 192

QY 175 GGSISTEYEEFGFVIGAYGA-----ADR--ITLQAPOLGNGKKEQMATGLKTD 221

Db 193 SYHAGLKYENAGFFGGVAGSPAKYADLNTDAERVAVNTANAHV---KDYQVRRVAYGTD 249

QY 222 ANNLY--LAANYGETRNPATPTNKFNTSGFANKTQDVLVLAQYQDFGLRPSIAYTKSK 279

Db 250 ANDLYSVAGQYERAKN-----NEVGSTGKKEHQVOVATAATAYRGG-NWTPRVSYAHGF 303

QY 280 AKDVEGIGVDLVN-----FEVATYFFNKNMSTYV 311

Db 304 KAKVNGVKD---ANYQYDOYIVGADYDFSKRTSALV 336

RESULT 12

US-08-798-760-6
; Sequence 6, Application US/08798760

; Patent No. 6013267

; GENERAL INFORMATION:

; APPLICANT: Blake, Milan S.

; APPLICANT: Tal, Joseph Y.

; APPLICANT: Qi, Hulin L.

; APPLICANT: Liang, Shu-Mei

; APPLICANT: Hronowski, Lucjan J.J.

; APPLICANT: Pullen, Jeffrey K.

; TITLE OF INVENTION: Method for the High Level Expression,

; TITLE OF INVENTION: Purification and Refolding of the Outer Membrane Group B

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/798,760

; FILING DATE: 11-FEB-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Esmond, Robert W.

; REGISTRATION NUMBER: 32,893

; REFERENCE/DOCKET NUMBER: 1438.0060002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 363 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-798-760-6

Query Match 8.7%; Score 157; DB 3; Length 363;

Best Local Similarity 24.7%; Pred. No. 6,7e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLYGKAVGLHYFSKNGENSYGNGDMTY-----ARLFGKGTQINSDLTGQW 61

Db 24 VTLYGKIKAGVEVSRKVDAGTYKRAQGGKSKTATQIADFGSKIGFGQEDLGNMKAIMWL 83

QY 62 EYNFGNNSGADAGQGNKRLAFAGLKYADVGSFDYGR--NNGVYVDALGYDMLPEFG 119

Db 84 E--QKASTAGTNSGMCN--ROSFILK---GGFGTVRAGNLTNVLKDSGDVNWAMESG 134

QY 120 GDTAVSDDFVGVGVAVIYRNSNFG--LVLDLNFVAVYGLK---NERDTARNSGDGV 174

Db 84 E---OKASIACTNSGWN--RQSFGLK---GGFGTVRAGNLTMTLKDSGDVNAMESG 134

QY 120 GDTAVSDDEFGRGVATYRNSNFFG--LVYDGLNFAVOYLK---NERDARRSNGDV 174

Db 135 SNT--EDVLGIGTIGRVSREISVRYDSPVAFSGSVQYVPRNANDVDYKKTKSSRE 192

QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNLOEAPYLGNGKKAQMATGLKYD 221

Db 193 SYHAGLKEKNAGFFGOYAGSPAKYADLNTDAERYAVNTANHPV---KDYVHRVYAGYD 249

QY 222 ANNTY--LAANYGETRNTPTTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSK 279

Db 250 ANDLYSVAGQYEAKN-----NEVGSTKGKKHETQVAAATPAATYRFG--NTPRVSYAHGF 303

QY 280 AKDVEGIGDVLVNY---FEVGATYYFNKNMSTYV 311

Db 304 KAKYNGVMD---ANYQYDQYLVGADYDSKRTSALV 336

RESULT 13

PCT-US94-08327-6

Sequence 6, Application PC/TUS9408327

GENERAL INFORMATION:

APPLICANT: The Rockefeller University

APPLICANT: 1230 York Avenue

APPLICANT: New York, New York 10021

APPLICANT: United States of America

APPLICANT: 12103 Indian Creek Court

APPLICANT: Beltsville, Maryland 20705

APPLICANT: United States of America

APPLICANT: Tal, Joseph Y.

APPLICANT: Q1, Huijin L.

APPLICANT: Hronowski, Lucjan J.

APPLICANT: Pullen, Jeffrey K.

TITLE OF INVENTION: Method for the High Level

TITLE OF INVENTION: Purification and Refolding of the Outer Membrane

TITLE OF INVENTION: Group B

TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Ave., Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/08327

FILING DATE: Herewith

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/096,182

FILING DATE: 23 July 1993

ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.

REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 1438, 006PC00

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 363 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-08327-6

Query Match 8.7%; Score 157; DB 5; Length 363;

Best Local Similarity 24.7%; Pred. No. 6,7e-07;

Matches 83; Conservative 45; Mismatches 150; Indels 58; Gaps 16;

QY 11 VDLYKAVGLHYFSKNGENSYGNGDMTY-----ARLKGKETQINSDLTYGQW 61

Db 24 VTLGYTIKAGVEVRVDACTYGAAGGSKSKATQIADFGSKIGKGQEDLNGKAIMQL 83

QY 62 EYNFGNNSGEGADQTNKPLAFAGLKYADVGSFDYGR--NYGVYDALGYTDLPRFG 119

Db 84 E---OKASIACTNSGWN--RQSFGLK---GGFGTVRAGNLTMTLKDSGDVNAMESG 134

QY 120 GDTAVSDDEFGRGVATYRNSNFFG--LVYDGLNFAVOYLK---NERDARRSNGDV 174

Db 135 SNT--EDVLGIGTIGRVSREISVRYDSPVAFSGSVQYVPRNANDVDYKKTKSSRE 192

QY 175 GGSISYEYEGFIVGAYGA-----ADR--TNLOEAPYLGNGKKAQMATGLKYD 221

Db 193 SYHAGLKEKNAGFFGOYAGSPAKYADLNTDAERYAVNTANHPV---KDYVHRVYAGYD 249

QY 222 ANNTY--LAANYGETRNTPTTNKFTNTSGFANKTQDVLVAOYQDFGLRPSIATYKSK 279

Db 250 ANDLYSVAGQYEAKN-----NEVGSTKGKKHETQVAAATPAATYRFG--NTPRVSYAHGF 303

QY 280 AKDVEGIGDVLVNY---FEVGATYYFNKNMSTYV 311

Db 304 KAKYNGVMD---ANYQYDQYLVGADYDSKRTSALV 336

RESULT 14

US-08-096-182A-2

Sequence 2, Application US/08096182A

Patent No. 5439808

GENERAL INFORMATION:

APPLICANT: Blake, Milan S.

APPLICANT: Tal, Joseph Y.

APPLICANT: Q1, Huijin L.

APPLICANT: Hronowski, Lucjan J.

APPLICANT: Pullen, Jeffrey K.

TITLE OF INVENTION: Method for the High Level Expression,

TITLE OF INVENTION: Purification and Refolding of the Outer Membrane

TITLE OF INVENTION: Porin Proteins from Neisseria Meningitidis

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Ave., Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/096,182A

FILING DATE: 23-JUL-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.

REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 1438, 0060000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 309 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-096-182A-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;
Best Local Similarity 25.1%; Pred. No. 4.1e-06;
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

35 NGDMTY-----ARLGFKEGTQINSDLTGQOMEYFNGGNSSEGADQTKRTL 83
19 NGQVTEVTATGIVDLSKIGFGQEDLNGLKAIWQVE---QKASIAIGTDSGWN--RQ 73
QY 84 AFAGLKYADVGSFDYGRNYGVYDALGYTDLPEFGD---TASDDEFVGRVGG---- 135
DB 74 SFTGLK-GGFGKLRVGRLSVLKDT-----GDINPMKSKDYLGVNKIAPEAR 121
QY 136 --VATYRNSNFEGLVDGLNFAVOYLGNKNERDARNSNGDVGSGISYEYEGFIVAGYA 193
DB 122 LISVRIDSPFEAGLSCSVQYAL-----NDNAGRHNSESYHAGFYKKGFEV--QYGG 172
QY 194 ADRTNLOEAQPLGNCK-KAEQWATGLKYDANNIYLAANGETRNATPINKFTNTSGFAN 252
DB 173 AKRHHQVOEGLEIEKYQIHRIVSG--YNDALY-ASVAVQODA-----KLTDA NSHN 224
QY 253 KTDVLLVAQYQDFGLRPSIATYKSKADVEG-IGDVLVNYFE--VGATYFENKNS 308
DB 225 SQTEVAATLAVRFG-NVTPRVSY---AHGFKGLVDADIGNEYDVVYGAEDYFSKRTS 279
QY 309 TYV 311
DB 280 ALV 282

RESULT 15

US-08-877-109-2
Sequence 2, Application US/08877109

Patent No. 5747287

GENERAL INFORMATION:

APPLICANT: Blake, Milan S.

APPLICANT: Tai, Joseph Y.

APPLICANT: Qi, Huijin L.

APPLICANT: Liang, Shu-Mei

APPLICANT: Hronowski, Lucjan J.J.

APPLICANT: Pullen, Jeffrey K.

TITLE OF INVENTION: Method for the High Level Expression,
Purification and Refolding of the Outer Membrane Group B

TITLE OF INVENTION: Porin Proteins from Neisseria meningitidis

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Ave., NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/877,109

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/431,264

FILING DATE: 28-APR-1995

APPLICATION NUMBER: 08/096,182

FILING DATE: 23-JUL-1993

ATTORNEY/AGENT INFORMATION:

NAME: Edmond, Robert W.

REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 1438.0060001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 309 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-877-109-2

Query Match 8.2%; Score 147.5; DB 1; Length 309;
Best Local Similarity 25.1%; Pred. No. 4.1e-06;
Matches 76; Conservative 46; Mismatches 116; Indels 65; Gaps 17;

35 NGDMTY-----ARLGFKEGTQINSDLTGQOMEYFNGGNSSEGADQTKRTL 83
19 NGQVTEVTATGIVDLSKIGFGQEDLNGLKAIWQVE---QKASIAIGTDSGWN--RQ 73
QY 84 AFAGLKYADVGSFDYGRNYGVYDALGYTDLPEFGD---TASDDEFVGRVGG---- 135
DB 74 SFTGLK-GGFGKLRVGRLSVLKDT-----GDINPMKSKDYLGVNKIAPEAR 121
QY 136 --VATYRNSNFEGLVDGLNFAVOYLGNKNERDARNSNGDVGSGISYEYEGFIVAGYA 193
DB 122 LISVRIDSPFEAGLSCSVQYAL-----NDNAGRHNSESYHAGFYKKGFEV--QYGG 172
QY 194 ADRTNLOEAQPLGNCK-KAEQWATGLKYDANNIYLAANGETRNATPINKFTNTSGFAN 252
DB 173 AKRHHQVOEGLEIEKYQIHRIVSG--YNDALY-ASVAVQODA-----KLTDA NSHN 224
QY 253 KTDVLLVAQYQDFGLRPSIATYKSKADVEG-IGDVLVNYFE--VGATYFENKNS 308
DB 225 SQTEVAATLAVRFG-NVTPRVSY---AHGFKGLVDADIGNEYDVVYGAEDYFSKRTS 279
QY 309 TYV 311
DB 280 ALV 282

Search completed: July 3, 2001, 14:56:46
Job time: 536 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:59:45 ; Search time 63.58 Seconds
(without alignments)
408,549 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEINRKNQKVDLYGKAVGL.....NKLGVGSDDTVAAGIVYQFA 341

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

219241

Minimum DB seq length: 0

219241

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	MMECF
2	1789	98.9	362	2	B85619
3	1113	61.5	351	2	B85513
4	1110	61.4	351	1	MMECFE
5	1070	59.1	351	2	S25520
6	1058	58.5	350	2	S25525
7	1051	58.1	365	1	MMBP2
8	1051	58.1	367	1	MMECP2
9	1044	57.7	350	2	S36606
10	1042	57.6	365	1	MMECP2
11	1035	57.2	377	2	D64888
12	1033	57.1	367	2	B85861
13	1030	56.9	363	2	S51104
14	1029.5	56.9	363	2	S43159
15	1023.5	56.6	378	1	MMEBP2
16	1000.5	55.3	378	1	AS9139
17	998.5	55.2	366	2	B85729
18	976.5	54.0	394	2	S34263
19	885	48.9	369	2	S70847
20	770	42.6	315	2	T46993
21	635.5	35.1	255	2	D85740
22	573.5	31.7	382	2	G84971
23	539.5	29.8	224	2	F85814
24	486	26.9	171	2	H64960
25	415	22.9	191	2	H64960
26	372.5	20.6	123	2	C85740
27	351	19.4	134	2	B64961
28	236.5	13.1	350	2	A82399
29	233.5	12.9	96	2	C64900

30	224	12.4	331	2	E82389	probable outer mem
31	197.5	10.9	330	2	PC4425	lectin-like adhesi
32	164.5	9.1	97	2	A57983	outer membrane por
33	157	8.7	360	2	S17456	outer membrane por
34	157	8.7	361	2	JC6314	major porin protei
35	155.5	8.6	50	2	T46990	hypothetical prote
36	153.5	8.5	312	2	PH0224	class-3 porin prot
37	153.5	8.5	371	2	S68072	major outer membra
38	151.5	8.4	371	2	S68072	major outer membra
39	150.5	8.3	313	2	JH0254	class-3 porin prot
40	150.5	8.3	331	2	S21407	major outer membra
41	150.5	8.3	390	2	I39570	probable porin (s1
42	149	8.2	323	2	S25514	outer membrane por
43	147.5	8.2	295	2	S25507	outer membrane pro
44	144.5	8.0	295	2	S25506	outer membrane pro
45	139.5	7.7	359	2	D64050	major outer membra

ALIGNMENTS

RESULT 1

MMECF
outer membrane porin ompF precursor - Escherichia coli
N:Alternate names: outer membrane protein 1a; outer membrane protein b
C:Species: Escherichia coli
C>Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 16-Jul-1999
C:Accession: A93449; A91301; A90314; A25029; H64832; A03431
R:Inokuchi, K.; Mutoh, N.; Matsuyama, S.; Mizushima, S.
Nucleic Acids Res. 10, 6957-6968, 1982

A:Title: Primary structure of the ompF gene that codes for a major outer membrane pro
A:Reference number: A93449; WUID:83050452

A:Accession: A93449

A:Molecule type: DNA

A:Residues: 1-362 <INOS>

A:Cross-references: GB:J01655; GB:M10311; GB:M10312; NID:9147009; PIDN:AAA24244.1; PI
A:Experimental source: strain K12

R:Mutoh, N.; Inokuchi, K.; Mizushima, S.
FEBS Lett. 137, 171-174, 1982

A:Title: Amino acid sequence of the signal peptide of OmpF, a major outer membrane pr
A:Reference number: A91301; WUID:82139379

A:Accession: A91301

A:Molecule type: DNA

A:Residues: 1-37 <MDT>

R:Chen, R.; Kramer, C.; Schmidmayr, W.; Chen-Schmeisser, U.; Henning, U.
Biochem. J. 203, 33-43, 1982

A:Title: Primary structure of major outer-membrane protein I (ompF protein, porin) of
A:Reference number: A90314; WUID:82256494

A:Accession: A90314

A:Molecule type: protein

A:Residues: 23-87, 'E', 89-138, 'G', 140-283, 'L', 285-362 <CHE>

R:Nogami, T.; Mizuno, T.; Mizushima, S.
J. Bacteriol. 164, 797-801, 1985

A:Title: Construction of a series of ompF-ompC chimeric genes by in vivo homologous r
A:Reference number: A91809; WUID:86033642

A:Accession: A25029

A:Molecule type: DNA

A:Residues: 33-63 <NOC>

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; WUID:97426617

A:Accession: H64832

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-362 <BAT>

A:Cross-references: GB:AE000195; GB:U00096; NID:91787156; PIDN:AACT4015.1; PID:917871
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: ompF; tolF; cmfB; coa; cry

A:Map position: 21 min

C:Complex: homotrimer

C:Function: <POR>
A:Description: forms aqueous channels that facilitate diffusion of small hydrophilic molecules
C:Function: <REC>
A:Description: receptor for bacteriophage T2
C:Superfamily: outer membrane protein phOE
C:Keywords: membrane protein; porin; trimer
F:1-32/Domain: signal sequence #status predicted <SIG>
F:32-362/Product: outer membrane porin ompF #status experimental <MAT>

Query Match	Score	DB 1	Length
99.8%	1805		362

Qy	1	AEIYNKOGNKVDIVGKAAGLHYHFSKNGENSYGGGMDTYARLGFQGTQINSDLTGGQ	60
Db	23	AEIYNKDKNKDVLGKAVGLHYHFSKNGENSYGGGMDTYARLGFQGTQINSDLTGGQ	82
Qy	61	WEYNFQGNNSGADAOQTGNKTRLAFAGLKYADVGSSEDIYGRNYGVYDALGTTDMLEPGG	120
Db	83	WEYNFQGNNSGADAOQTGNKTRLAFAGLKYADVGSSEDIYGRNYGVYDALGTTDMLEPGG	142
Qy	121	DTAYSDDFEYVGVGAVATYRNSNFEGLVDGLNEFAVOYLGKNERDTPARRSNGDVGGSISY	180
Db	143	DTAYSDDFEYVGVGAVATYRNSNFEGLVDGLNEFAVOYLGKNERDTPARRSNGDVGGSISY	202
Qy	181	EYEGFGIVAGYCAADRNLQEAQPLGNGKKAQWATTKLYDANNITLYLANGETNNATPI	240
Db	203	EYEGFGIVAGYCAADRNLQEAQPLGNGKKAQWATTKLYDANNITLYLANGETNNATPI	262
Qy	241	TKKFNTSGFAKKTODVLLVAOYQDFGLRPSIATYKSKAKDVEIGDVLNYPEVGAT	300
Db	263	TKKFNTSGFAKKTODVLLVAOYQDFGLRPSIATYKSKAKDVEIGDVLNYPEVGAT	322
Qy	301	YFENKRMSTYYDYIINQIDSDNKLGVGSDPYAVAGLYVOF	340
Db	323	YFENKRMSTYYDYIINQIDSDNKLGVGSDPYAVAGLYVOF	362

RESULT 2
B85619
outer membrane protein Ia (Ia,b,F) [Imported] - Escherichia coli (strain O157:H7)

C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85619
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: B85619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <STO>
A:Cross-references: GB:AE005174; NID:q12514104; PIDN:AAG55414.1; GSPDB:GN00145; UMGF:Z171
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ompF
C:Superfamily: outer membrane protein phoE

Query Match	98.98;	Score 1789;	DB 2;	Length 362;
Best Local Similarity	99.48;	Pred. No. 4.3e-122;		
Matches 338; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	1	AETLYKDKGKVDLYKRAVGLHYHESKNGNENYGGNGDMVTARLGFKEETQINSNDLGVGQ	60
Db	23	AETLYKDKGKVDLYKRAVGLHYHESKNGNENYGGNGDMVTARLGFKEETQINSNDLGVGQ	82
Qy	61	WEYNFGNNSGADQOTGNKTRLAFLAGLTAVDVSFDYGRNYGVYDALGYTDLPEFGG	120
Db	83	WEYNFGNNSGADQOTGNKTRLAFLAGLTAVDVSFDYGRNYGVYDALGYTDLPEFGG	142

QY	121	DTVASDPEFVGRGVGATVYTRNSNFFGLVNDGINFAYOVYLGKNEBRTARSGDGGVGSISY	180
Db	143	DTVASDPEFVGRGVGATVYTRNSNFFGLVNDGINFAYOVYLGKNEBRTARSGDGGVGSISY	202
QY	181	EYEFEGVIGAVGADPRNLEBAPCLGNGKKAEMQWATLKYDANNIYLAANYGETRNPPI	240
Db	203	EYEFEGVIGAVGADPRNLEBAPCLGNGKKAEMQWATLKYDANNIYLAANYGETRNPPI	262
QY	241	TNKEPTNTSGFANKTQDVLVLAQYQDFGLRPSIAYYTKSKAKDVEGIGDVLVNYFEYVAT	300
Db	263	TNKEPTNTSGFANKTQDVLVLAQYQDFGLRPSIAYYTKSKAKDVEGIGDVLVNYFEYVAT	322
QY	301	YFENKMSYTYDYIINQIDSDNKLGVGSDDTVAAGIYYQF	340
Db	323	YFENKMSYTYDYIINQIDSDNKLGVGSDDTVAAGIYYQF	362

```

RESULT      3
B85513
outer membrane pore protein E (E Ic, NmpAB) [imported] - Escherichia coli (strain O157
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85513
R:Perena, N.W.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: B85480; MUID:21074935; PMID:11206551
A:Accession: B85513
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-751 <570>
A:Cross-references: GB:AB005174; NTD:G12513015; PIDN:AAG54566.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: phoE
C:Superfamily: outer membrane protein phoE

```

Query Match	61.5%	Score 1113;	DB 2;	Length 351;
Best Local Similarity	62.1%;	Pred. No. 2.9e-73;		
Matches 213; Conservative	38;	Mismatches 76;	Indels 16;	Gaps 5

```

QY      1 AEWKQCNKYVDLGVKXGKXVHLFEKSGKNGNSYGGNGMGTARLGEKFEETDINSDLGVGQ 60
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      22 AETTKDKNKLIDVGRKAKAMHYMSDNDK -----DGQSYIRGFKFETDINOQLGVGR 76

QY      61 WEYNQGNNSGADAOQTGNKTRLAPAGLKYADVGSFEDYGRNYGVYTDALCYTHMLPEFGG 120
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      77 WEAFEGAKKAESDTPAO -QKTRLPAGLKYKDLGSPFYGRMLGALYVEAMTDMPEFGG 134

QY      121 D-TATSDQFPGVGRGVATYRNSNPFGLVDLGNVAOYLKGNENHDPTRRSGNGGVGSGIS 179
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      135 DSSAOTDNEMTKRASGLTRTYRTEDFEYIIDLNLTLQYQKNEHROVKKQNGGFGTSLT 194

QY      180 YELEG -FGIVGAGADRTMLQEHQPLGNGKKAEBQWATGTLKDANNIYLAANYGEFRMA 237
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      195 YPFGSDEFAISGAYNSRDTBEOINQSGTKRABAMATGTLKDANNIYLATFVSETRKM 254

QY      238 FDTITKFTNTSGFAKTDQDVLVAOYQDFGLRARSIAVTKSKAKDVEIGIDVDLVNVEF 297
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      255 FFL-----TGFAKKTQFEVAOYQDFGLRSLSLGIVLTKSGKDIDGIDEDLVNVIDV 308

QY      298 GATYYFNKNMSTYYDIILNQIDSDNKLGVGSDPTVAAGIYQF 340
        | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      309 GATYYFNKNMSAFDYDKINQDSDNKLININDDIIVAGMTQF 351

```

```

RESULT      4
MMECP2
outer membrane porin phoE precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 13-Jun-1993 #sequence_revision 13-Jun-1993 #text_change 21-Jul-2000

```


A:Gene: phoE
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-330/Product: outer membrane porin phoE #status predicted <MAT>

Query Match	58.5%	Score 1058	DB 2	Length 350
Best Local Similarity	58.0%	Pred. No. 2.8e-69		
Matches 199	Conservative 46	Mismatches 82	Indels 16	Gaps 5

DQ 1 AEINKDGANKVDLYGKAVGLHFFSKNGENSGYGGNDMTYARLGFKEEQINSDLTGYGQ 60
||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
DQ 21 AEVYNKNGNKLDVYGKVAKAMHYMSDYDSK----DDQSIVREGEFGKETQINDLTGYGR 75

Qy 61 WEINFGGNSSEGADQGTGNKRLAFAGLKYADGSGFDYGRNRYVYDALGSTDMLTEFGG 120
 || | : : | | | | | | | : | | | | | | : | | | | | |
 Db 76 WEAEFASNAKAESSDSS--QKTRLAFAGLKLKDIGSFDYGRNLGALYDVEAWTDMPEPEFGG 133

Df **Qy**

121 D-PAVSDDEFGVRGGGVAITYRNSNFELVDGLGFAYVOITLGNERNRTARRSGDGVGSIS 179
| : | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
134 DSAQTDFNFWTKRASGLATYKNTDFEGIVDGDLTLTYQGKNEDRDVKKKQNGDGFGGISVS 193

180 IEIEG-FEGVIGAGGAADRIQLQEAQPLGNGKAEQWAGLKYANNILYLANIGETRNA 237
 194 YDEGSGDFAVSGAYTILSDRTREQNLQRGTGDAAEAWATGVKTYDANDIYIATFYSEIRNM 253

```
Dc      254 TPV-----SGGFANKTQNFEAIVIQQFDFGLRPSLGYVLKSKDIEGVGSDELVNVIIDV    307
```

308 GATYYENKMSAEVFDIKINQDSDNTLGINDDDVAIGLYQF 3500

RESULT 7
MMBPP2
outer membrane porin lc precursor - phage PA2

C:\Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 02-Jul-1998
C:\Accession: D25647
R, Blasband, A.J.; Marcotte Jr., W.R.; Schnatman, C.A.

A:Title: Structure of the lc and nmhc outer membrane porin protein genes of lambdoid bacteriophages
A:Reference number: A25647; MUID:86304457
A:Accession: D25647

```

A:Molecule type: DNA
A:Residues: 1-365 <BLA>
C:Genetics:
A:Gene: lc
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
E:1.23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

```

Query Match	58.1%	Score 1051	DB 1	Length 365
Best Local Similarity	59.8%	Pred. No. 9.3e-69		
Matches 211	Conservative 35	Mismatches 83	Indels 24	Gaps 7

```
QY 1 AEIYNKDGKAVDLYGKAVGLHYFSKGGNGENSYGNGDMHYARLGEKGETQINSDLTGCGQ 60
      ||||| :||:||||| ||| : : ||| ||||| ||||| |||: ||
Db 24 AEIYNKDSNKLDTGKAVNAKHVFSSNDAD-----DGDITYARLGEKGETQINDQLTGFGQ 78
```

```
Qy      61 WEINFGNNSEGADQTGNTKRLAFLAGLKYADVGSFDGRNRYGVYDALGTMDMLPEEGG 120
        ||| :||| : : : ||||||| : | | ||||| || :||| |||||
Db      79 WEYEFFKGNDAE-SGGSSSKDKTRLAFLAGLKFGDYGSDYGRNRYGVAYVDIGAWTDVLP EEG 137
```

DQ 121 D^T-A YSDDFVGRGCVATYRNSNFGGLVDGLNEFAVQLTGKNER---DTRARSNNGDVGG 176
 || :||| || | |||||:||||||| || ||||| ||
 Db 138 DWTQTVDVEMTGRITTFGATYRNNDFFGLVDGLNEFAAQYGKNDRSDFDNYTEGNCGDGGEF 197

```

QY 177 SISYEYGFVIGVAYGAADPTNOE-----ADPLGKKAEOAWGLKRDANNITLAAN 230
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 SATYEYGFVIGATYAKSDPTDVOVNAKYLPEYFASGKAEBWAGLKDANNITLATT 237
QY 231 YGETRNATPTLNKPTNTSGFANKTODVLVAOYOFDEGLSPSIATYTRSKAKDVEGIGDVD 290
      | | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 258 YSEFQNTVFRADHV-----ANKAQNREAVAOYQOFDEGLSPSAYTYLQSKKRDLCGWGDD 312
QY 291 LVNTEYGVATYYFNKNKMSVYDIINQISDN--KLGVSDDTVAVGIVYOF 340
      | | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 LVKXVVDVGATYYFNKNKMSFVYDKINILDKNDPKALGVSTDDIVAVGLVYOF 365

```

```

      RESULT      8
      MMECCPC
      outer membrane porin omcB precursor - Escherichia coli

```

C/Species: *Escherichia coli*
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1989
C/Accession: A20867; A18885; B25029; E64991
C/Mizuno, T.; Chou, M.Y.; Inouye, M.

U. Biol. Chem. 258, 932-940, 1983
A:Title: A comparative study on the genes for three porins of the *Escherichia coli* outer membrane
A:Reference number: A20867; MUID:83213433
A:Accession: A20867

A;Residues:1-367<MIZ>
A;Cross-References:GB:K00541; GB:M10314; GB:M14188; NID:9147007; PIDN:AAA24243.1; PIR:Q00000
R;Mizuno, T.; Chou, M.Y.; Inouye, M.

A:Title: DNA sequence of the promoter region of the ompC gene and the amino acid sequ
A:Reference number: A18885; MUID:83132326
A:Accession: A1885

A;Residues: 1-40 <MI2>
R;Nogami, T.; Mizuno, T.; Mizushima, S.
J. Bacteriol. 164, 797-801, 1985

A;Reference number: A91809; MUID:86033642
A;Accession: B25029
A;Molecule type: DNA
A;Postdoc: 32-57 <NOC>

A.; Rose, D. J.; Mau, B.; Shao, Y. R.; Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M. Science 277, 1453-1462, 1997

A:Reference number: Ab4720; MUID:9/42661/
A:Accession: E64991
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-36 / <BLAST>
A:Cross-references: GB:AE000310; GB:U00096; NID:92367131; PIDN:AMC75275.1; PID:917685
A:Experimental source: strain K-12, substrain MG1655
A:Genetics:

```
A;scene: compc
A;Map position: 47 mir
C;Function:
A;Description: one of
```

C;Keywords: membrane protein; porin; trimer
F.1-21/Domain: signal sequence #status predicted <SIG>
F.22-367/Product: outer membrane porin ompC #status predicted <MAT>

Query Match	Score	DB 1	Length
Best Local Similarity	58.1%;	1051;	367;
	58.9%;	Pred. No. 9.4e-69;	

```

QY      1 AEIYNKDKNKVDLYGKAVGLHYFSKNGENSXGNGDMTVARLGFKGETQINSIDLTYGYG 600
      ||:|||||:||||| ||||| : ||||| |||||: |||||
Db     22 AEIYNKDKNKDLIYGKVDGLHYFS-----DNKDVSDQTYMLGFKGEQVTDQLTYGYG 760

```

Db 22 AEYYNKDGKRLDLGKVDGLHYFS-----DNKDDVDGDQTYMRLGFKGEQVTTDQLLGGQ 76

Qy 61 WEYFQGNNSGADAQTGNKTRLAFLAAGLKYADVSGFDYGRNYGVVYDALGYTDMLEPFGG 120

```

Db 77 WEYQIOGNSEN---ENNSTRAFAAGLKFQDVGSPDYGNKYVYDVYVSWTVLPEFGG 133
Qy 121 DPAVSDDFEVGRVGVATYRNSNFGVLVDGLNFAVOYLGRNERDT-----A 166
Db 134 DTVGSNFMQORNGRNGATYRNTDFPGVLVDGLNFAVOYQKNGNSGCGFTSGVTNNGRDA 193
Qy 167 RNSNGDVGGSISYEYEGGIGVAGADRTNIOE-AQPLGSKKAQOMATGLKYDANNI 225
Db 194 LRONGDVGGSISTYDEYEGGIGCAISSKRTDAONTAAAYIGNDRAETETGYGLKYDANNI 253
Qy 226 YLAANQGERNATPITNKFTNTSGFANKQDVLVAOYQDFGLRPSIATKSKARDV-E 284
Db 254 YLAQYTOTYNNARV-----GSLGMANKAQNFEFAVOYQDFGLRPSLAYLQSKGNLGR 308
Qy 285 GIGDVLVNVFEVATYFNKNSSTVVDYIINDSDNKL-----GVGSDDTVAVGIIYQF 340
Db 309 GYVDEDLKTVYDVGATYFNKNSSTVVDYKINLLD-DNQRTDAGINTDITVALGLVYQF 367

RESULT 9
outer membrane porin phoE precursor - Salmonella typhi
C:Species: Salmonella typhi
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S36606
R:Torres, A.; Puente, J.L.; Calva, E.
submitted to the EMBL Data Library, August 1993
A:Reference number: S36606
A:Accession: S36606
A:Molecule type: DNA
A:Residues: 1-350 <OR>
A:Cross-references: EMBL:X74595; NID:g397159; PIDN:CA53672.1; PID:g397160
A:Experimental source: Isolate IMS-1
A:Note: this is one of the proteins induced when the E. coli cells are grown under phospho-
d compounds, and some other negatively charged solutes
C:Genetics:
A:Gene: phoE
C:Complex: homotrimer
A:Description: allows passive diffusion of small hydrophilic molecules with molecular we
A:Note: Induced under phosphate limitation
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-350/Product: outer membrane porin phoE #status predicted <OPP>

Query Match 57.7%; Score 1044; DB 2; Length 350;
Best Local Similarity 57.4%; Pred. No. 2.8e-68;
Matches 197; Conservative 47; Mismatches 83; Indels 16; Gaps 5;

Qy 1 AEIYNKDNKNDYLYGKAVGLHYFSKNGSENSYSGNGMTYARLGFGEINOINSLDLYGQ 60
Db 21 AEIYNKNGKLDYLYGKAVKAMHYSDYDSK-----DGDQSYVRFEGFGKTIINOQLTYGR 75
Qy 61 WEYNFOGNNSEGADQGTGKTRLAFAGLKAYADVSFDYGRNRYGVYDALCYTDLPEFGG 120
Db 76 WEYEGAGNKAESDSQ--QKTRLAFAGLKLDIGSFDYGRNLGALYDEAMTDFPEFGG 133
Qy 121 D-TYASDDFEVGRVGVATYRNSNFGVLVDGLNFAVOYLGRNERDTRARRNGSGVGSIS 179
Db 134 DSSAQDNTMTKRAASGLATYRNTDFGIVDGLDTLOYQKKNDRDKKONGSGGFGTSVS 193
Qy 180 YEYEG--FCIVGAYGAADRTNIOEAPLGNKKAQOMATGLKYDANNIYLAANYGETRNA 237
Db 194 YDGGSDPAVSAGYATSDTRREONLQRGTDGAEGHATCVKTDADITYATYSETRNA 253
Qy 238 TPIITNKFTNTSGFANKQDVLVAOYQDFGLRPSIATYTSKAKADVEGIGDVLVNVFEV 297
Db 254 TPV-----SGFRANKTQNEAVIYOQDFGLRPSLGYVLSKSKDIEGVSEDLVNVYIDV 307
Qy 298 GATYFNKNSSTVVDYIINDSDNKLGVGSDDTVAVGIIYQF 340

```

```

Db 308 GATYFNKNSAEDYKTIINDSDNTGINDDITVALGIIYQF 350

RESULT 10
outer membrane porin mmpC precursor - Escherichia coli cryptic lambdaoid prophage DLP1
MECNC
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: A25647; S66594; G64787
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaltman, C.A.
J. Biol. Chem. 261, 12723-12732, 1986
A:Title: Structure of the Ic and mmpC outer membrane protein genes of lambdaoid
A:Reference number: A25647; MUID:86304457
A:Contents: mutant strain CS384
A:Accession: A25647
A:Molecule type: DNA
A:Residues: 1-365 <BA>
R:Mandl, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.
J. Mol. Biol. 257, 561-573, 1996
A:Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia
A:Reference number: S66579; MUID:96196428
A:Accession: S66594
A:Molecule type: DNA
A:Residues: 347-365 <MAH>
A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CA63325.1; PID:g1051145
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64787
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'MNIRAYATSEFNNSKKGLT', 1-325, 'N', 327-347, 'EGANKSLI' <BLAT>
A:Cross-references: GB:U000160; GB:U00096; NID:g178651; PIDN:AC03654.1; PID:g17867
A:Experimental source: strain K-12, substrain MG1655
A:Comment: In wild-type strains of E. coli K-12, the mmpC open reading frame is inter-
ain CS348, the ISS element has been deleted and mmpC is expressed.
C:Genetics:
A:Gene: mmpC
A:Map position: 12 min
A:Genome: cryptic lambdaoid prophage DLP12
C:Superfamily: outer membrane protein phoE
C:Keywords: membrane protein; porin; trimer
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-365/Product: outer membrane porin mmpC #status predicted <MAT>

Query Match 57.6%; Score 1042; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 4.2e-68;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

Qy 1 AEIYNKDNKNDYLYGKAVGLHYFSKNGSENSYSGNGMTYARLGFGEINOINSLDLYGQ 60
Db 24 AEIYNKDSNKLDTLYGKANAKHFSSNDAD-----DGYTTRARLAFKQETIINOQLTGFQ 78
Qy 61 WEYNFOGNNSEGADQGTGKTRLAFAGLKAYADVSFDYGRNRYGVYDALCYTDLPEFGG 120
Db 79 WEYEFKGNRAE-SGSSKDKTRLAFAGLKRGDYSIDYGRNRYGVYDALCYTDLPEFGG 137
Qy 121 DT-AYSDDFEVGRVGVATYRNSNFGVLVDGLNFAVOYLGRNERD--DTARRNSGDSVG 176
Db 138 DFTWOTDVFMTORATGATYRNNDFEGLVDGLNFAVOYQKKNDRSDPDNTTEGNGDGF 197
Qy 177 SISYEYEGGIVAGYGAADRTNIOE-----AQPLGKKAQOMATGLKYDANNIYLAAN 230
Db 198 SATYEYEGGIVAGYKSDTDTQVNAAGVLPYFASGKAHEWAAGLKYDANNIYLAAT 257
Qy 231 YGETRNATPITNKFTNTSGFANKQDVLVAOYQDFGLRPSIATYTSKAKADVEGIGDVL 290
Db 258 YSETOKNTVFADHFV-----ANKAQNFEVAVOYQDFGLRPSVAIYLSKSKDLGVWGDD 312

```


OY 61 WEYNFQGNNSGADQGTGKRLAFAGLKYADVGSFDYGRNYYVDALGYTDLPERGG 120
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 77 WEYVQANNTSSSDQAM--TRLAFAGLKEGSDAGSFDYGRNYYVDTSMTDVLPERGG 134
 OY 121 DTAVSDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNK----ERDT-----ARRSN 170
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 135 DTYSQDNFLGRANGAVATYRNSDFGLVDGLNFAVOYQGNKGSVSGEATNNGALKON 194
 OY 171 GDVGGISIEY-YEGFGIVAGYGAADRTNLOEAPLGNKKAEDQWATGLKYDANNITAA 229
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 195 GDGFGTSTYVDIFDGISAGFAVANSKRITDQNOQLLEGDHAETITGLKYDANNITLAT 254
 OY 230 NYGTRNATPTTNKFTMTSGFANKTODVLVAOYQFDGLRPSIATYRSKAKDVEGCDV 289
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 255 QYQTYNATR-----AGSLGRANKAQNFEVAAOYQFDGLRPSVAYLQSKGDLNGYDQ 309
 OY 290 DLVNYFEVATYFNKMNSTYVDYIIQIDSDN--KLGVSDDTVAAGIYQF 340
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 310 DLKYYVGVATYFNKMNSTYVDYKINLIDNSFTRSKGISTDVVALGLIYQF 363

RESULT 14

S43159

outer membrane porin ompF precursor - Salmonella typhimurium

C:Species: Salmonella typhimurium

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999

C:Accession: S43159

R:Venegas, A.; Gomez, I.; Bruce, E.; Martinez, M.

submitted to the EMBL Data Library, March 1994

A:Description: PCR amplification and cloning of the Salmonella typhimurium ompF porin 96

A:Reference number: S43159

A:Accession: S43159

A:Molecule type: DNA

A:Residues: 1-363 <VEN>

A:Cross-references: EMBL:Z31594; NID:9468740; PIDN:CAA83471.1; PID:9468741

C:Genetics:

A:Gene: ompF

C:Superfamily: outer membrane protein phoE

C:Keywords: membrane protein; porin; trimer

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-363/Product: outer membrane porin ompF #status predicted <MAT>

Query Match 56.98; Score 1029.5; DB 2; Length 363;
 Best Local Similarity 57.68; Pred. No. 3.3e-67;
 Matches 200; Conservative 54; Mismatches 80; Indels 13; Gaps 8;

OY 1 AEIYNKQGNKYDLGKAVGLHYFSKNGENSYGNGDMTYARLGFKEGTQINSDLTYGQ 60
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 23 AEIYNKQGNKLDLGGKAVGRHVWTT--TGDSK--NADQYAOIGFKEGTQINTDLTGFGQ 78
 OY 61 WEYNFQGNNSGADQGTGKRLAFAGLKYADVGSFDYGRNYYVDALGYTDLPERGG 120
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 79 WEYTKADNRAEG--EQGNSNLVRLAFAGLKYAEVGSIDGRNYYGYVDVSTIDMAPYRSG 137
 OY 121 DT---AVSDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNKNERDTARRS--NGDVG 176
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 138 ETWGAATDNTMTSRAGLLTYRNSDFGLVDGLSFQIOYQGNKDNHSINSQMDGVGY 197
 OY 177 SISVEYFGFVGVAGYGAADRTNLOEAPLGNKKAEDQWATGLKYDANNITAAVANGERN 236
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 198 TMALEFGDFGTAAVYSNKRNDQDDRD--GNGDRASAAVAKTDANNVYLAAYAEIRN 256
 OY 237 ATPITNFTNTSGFANKTODVLVAOYQFDGLRPSIATYRSKAKDVEGCI--GDVLDVNYF 295
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 257 MSIVENTYTDVEANKTONLEVAAOYQFDGLRPSIATYRSKAKDVEGCI--GDVLDVNYF 316
 OY 296 EVGATYFNKMNSTYVDYIIQIDSDNKL--VGSDDTVAAGIYQF 340
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 317 QAGATYFNKMNSTYVDYIIQIDSDNKL--VGSDDTVAAGIYQF 363

RESULT 15

MMBPC

outer membrane porin ompC precursor - Salmonella typhi

C:Species: Salmonella typhi

C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C:Accession: JQ0119; S01248; S10122

R:Punette, J.L.; Alvarez-Scherer, V.; Gosset, G.; Calva, E.

Gene 83, 197-206, 1989

A:Title: Comparative analysis of the Salmonella typhi and Escherichia coli ompC genes

A:Reference number: JQ0119; MUID:90060831

A:Accession: JQ0119

A:Molecule type: DNA

A:Residues: 1-378 <PUE>

A:Cross-references: GB:M1424; NID:9154207; PIDN:AAA27169.1; PID:9154208

R:Venegas, A.; Gomez, I.; Zaror, I.; Yudelevich, A.

Nucleic Acids Res. 16, 7721, 1988

A:Title: The nucleotide sequence of the Salmonella typhi ompC porin gene.

A:Reference number: S01248; MUID:88319959

A:Accession: S01248

A:Molecule type: DNA

A:Residues: 1-5, '0', 7-361, 363-378 <VEN1>

A:Cross-references: EMBL:X07835; NID:947796; PIDN:CAA30688.1; PID:947797

A:Experimental source: strain Ty2

R:Venegas, A.

submitted to the EMBL Data Library, May 1988

A:Reference number: S10122

A:Accession: S10122

A:Molecule type: DNA

A:Residues: 1-361, 363-378 <VEN2>

A:Cross-references: EMBL:X07835; NID:947796; PIDN:CAA30688.1; PID:947797

C:Genetics:

A:Gene: ompC

A:Map position: 45 min

C:Superfamily: outer membrane protein phoE

C:Keywords: membrane protein; porin; trimer

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-378/Product: outer membrane porin ompC #status predicted <MAT>

Query Match 56.68; Score 1023.5; DB 1; Length 378;
 Best Local Similarity 57.38; Pred. No. 9.4e-67;
 Matches 209; Conservative 44; Mismatches 79; Indels 33; Gaps 10;

OY 1 AEIYNKQGNKYDLGKAVGLHYFSKNGENSYGNGDMTYARLGFKEGTQINSDLTYGQ 60
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 22 AEIYNKQGNKLDLGGKAVGLHYFS----DDKSDGDDTYMRIGFKEGTQVNDLTYGQ 76
 OY 61 WEYNFQGNNSGADQGTGKRLAFAGLKYADVGSFDYGRNYYVDALGYTDLPERGG 120
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 77 WEYIQGNQTEGSSN--DSWTRVAFAGLKFADAGSFDYGRNYYGYVDVSTIDMAPYRSG 133
 OY 121 DTAVSDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNK----ERDTAR--RSNGD 172
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 134 DTYGADNFMQGRNGATYATYRTDFGLVDGLDFALQYQGNKGSVSGEATNNGRSLNKG 193
 OY 173 GVGGSISIEY-EGEGIVATY----GAADRTNLOEAPLGNKKAEDQWATGLKYDANNITL 227
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 194 GYGGSILYAIIEGGSVGAITTSKRTADQNTANRLGNGDRAVYVYGGILKYDANNITL 253
 OY 228 AANGETRNATPI--TNKFTNTS---GFANKTODVLVAOYQFDGLRPSIATYRSKAKDV 283
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 254 AAQYSQITNATRFGTSNNSNPSTSGFANKAQNFEVAAOYQFDGLRPSVAYLQSKGDI 313
 OY 284 E-----GIGDVLVNYFEVATYFNKMNSTYVDYIIQIDSDN--KLGVSDDTVAAG 335
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 314 SNGGASVGDQDLYKYVDVGVATYFNKMNSTYVDYKINLIDNKDPTDAGINTDIDVALG 373
 OY 336 IYQF 340
 ||| ||| : | ||||| : ||||| ||||| : ||||| |||||
 Db 374 IYQF 378

Search completed: July 3, 2001, 14:59:46

Job time: 371 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:50 ; Search time 37.35 Seconds

(without alignments)
312.748 Million cell updates/sec

Title: US-09-490-291-9

Perfect score: 1809

Sequence: 1 AEIYNKGNVDLXGRAVGL.....NKLGVSDPIVAVGIVYQFA 341

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_39:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1805	99.8	362	1	OMP_F_ECOLI
2	1110	61.4	351	1	PHOE_ECOLI
3	1072.5	59.3	350	1	PHOE_ENTCL
4	1072	59.3	374	1	OMP_F_SBRMA
5	1070	59.1	351	1	PHOE_CITFR
6	1058	58.5	350	1	PHOE_SALTY
7	1051	58.1	367	1	OMP_C_ECOLI
8	1050	58.0	351	1	PHOE_KLEPN
9	1046	57.8	365	1	PHOE_BPPAZ
10	1044	57.7	350	1	PHOE_SALTY
11	1042	57.6	365	1	NMP_C_ECOLI
12	1041	57.5	349	1	PHOE_KLEOX
13	1035	57.2	377	1	OMP_E_ECOLI
14	1031.5	57.0	363	1	OMP_F_SALTY
15	1030	56.9	363	1	OMP_C_KLEPN
16	1028.5	56.9	363	1	OMP_F_SALTY
17	1027	56.8	383	1	OMP_F_SALTY
18	1023.5	56.6	378	1	OMP_C_SALTY
19	1000.5	55.3	378	1	OMP_C_SALTY
20	999	55.2	397	1	YEDS_ECOLI
21	976.5	54.0	394	1	OMP_S_SALTY
22	902	48.9	342	1	OMP_C_KLEPN
23	894.5	48.9	376	1	OMP_F_SALTY
24	885	48.9	369	1	OMP_F_SALTY
25	825.5	45.6	282	1	OMP_D_SALTY
26	573.5	31.7	382	1	OMP_U_BUCAT
27	236.5	13.1	341	1	OMP_U_VIBCH
28	218	12.1	341	1	OMP_L_PHOPR
29	170.5	9.4	353	1	OMP_2_HAEIN
30	151.5	8.4	371	1	OMP_2_HAEIN
31	150.5	8.3	331	1	OMP_NEIMB
32	141.5	7.8	360	1	OMP_2_HAEIN
33	139.5	7.7	359	1	OMP_2_HAEIN

34	139	7.7	361	1	OMP_2_HAEIN	048217 haemophilus
35	138.5	7.7	348	1	OMP_1_NEIGO	P18195 neisseria g
36	138.5	7.7	361	1	OMP_2_HAEIN	P20148 haemophilus
37	135.5	7.5	369	1	OMP_2_HAEIN	048218 haemophilus
38	133	7.4	1829	1	FRPC_NEIMB	0919V5 neisseria m
39	131.5	7.3	348	1	OMP_2_NEIGO	P20148 neisseria m
40	131	7.2	331	1	OMP_2_NEIMB	P30688 neisseria m
41	131	7.2	1829	1	FRPC_NEIMB	P55127 neisseria m
42	130	7.2	367	1	OMP_2_HAEIN	048219 haemophilus
43	129.5	7.2	363	1	OMP_2_HAEIN	048220 haemophilus
44	129	7.1	331	1	OMP_1_NEIMB	P30687 neisseria m
45	129	7.1	331	1	OMP_3_NEIMB	P30689 neisseria m

ALIGNMENTS

RESULT	ID	OMP_F_ECOLI	STANDARD:	PRT:	362 AA.
AC	P02931;				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	15-JUL-1999 (Rel. 38, Last annotation update)				
DE	OUTER MEMBRANE PROTEIN F PRECURSOR (PORIN OMPF) (OUTER MEMBRANE				
DE	PROTEIN 1A, IA, OR B).				
GN	OMP OR TOLP OR CMLB OR COA OR CRY.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RC	MEDLINE-83090452; PubMed-6294623;				
RA	Inokuchi K., Mutoh N., Matsuyama S.-I., Mizushima S.;				
RT	*Primary structure of the ompf gene that codes for a major outer				
RT	membrane protein of Escherichia coli K-12.;				
RL	Nucleic Acids Res. 10:6957-6968(1982).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12 / MG1655;				
RC	MEDLINE-97426617; PubMed-9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RT	Mau B., Shao Y.;				
RT	*The complete genome sequence of Escherichia coli K-12.;				
RL	Science 277:1453-1474(1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RC	MEDLINE-97061202; PubMed-8905232;				
RA	Oshima T., Alpha H., Baba T., Fujita K., Hayashi K., Honjo A.,				
RA	Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kasahiro K.,				
RA	Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,				
RA	Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,				
RA	Sampel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,				
RT	Yano M., Horikuchi T.;				
RT	*A 718-kb DNA sequence of the Escherichia coli K-12 genome				
RT	corresponding to the 12.7-28.0 min region on the linkage map.;				
RL	DNA Res. 3:137-155(1996).				
RN	[4]				
RP	SEQUENCE OF 1-37 FROM N.A.				
RC	MEDLINE-82139379; PubMed-7037455;				
RA	Mutoh N., Inokuchi K., Mizushima S.-I.;				
RT	*Amino acid sequence of the signal peptide of OmpF, a major outer				
RT	membrane protein of Escherichia coli.;				
RL	FEBS Lett. 137:171-174(1982).				
RN	[5]				
RP	SEQUENCE OF 23-362.				
RC	MEDLINE-82256494; PubMed-7049161;				
RA	Chen R., Kramer C., Schmidmayr W., Chen-Schmeisser U., Henning U.;				

Query Match 99.8%; Score 1805; DB 1; Length 362;
 Best Local Similarity 100.0%; Pred. No. 1.7e-121;
 Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEIYNKDGKNDVLYGKAVGLHFFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGXGQ 60
 |||||||
 DB 23 AEIYNKDGKNDVLYGKAVGLHFFSKNGENSGYSGNDMTYARLFGKGTQINSDLTGXGQ 82
 |||||||
 QY 61 WEYNFQGNNEBGADQGTGKTRFLAFLAKTADVGSFDYGRNRYVYDALGYTDMLEPFGG 120
 |||||||
 DB 83 WEYNFQGNNEBGADQGTGKTRFLAFLAKTADVGSFDYGRNRYVYDALGYTDMLEPFGG 142
 |||||||
 QY 121 DTYVSDPEFVGRGVATYRNSNFEGLVDGNFAVOYLGNKERTARNSDGGSGSY 180
 |||||||
 DB 143 DTYVSDPEFVGRGVATYRNSNFEGLVDGNFAVOYLGNKERTARNSDGGSGSY 202
 |||||||
 QY 181 EYEGFGIVAGVAGADRNTLGEAOPLGNGKRAEQMATGLKTDANNIYLAANGETRNPPI 240
 |||||||
 DB 203 EYEGFGIVAGVAGADRNTLGEAOPLGNGKRAEQMATGLKTDANNIYLAANGETRNPPI 262
 |||||||
 QY 241 TNKFTNTSGFRANKTQDVLVAQVDFGLRPSIAYTKSKAKDVGIGDVLVNFVGAT 300
 |||||||
 DB 263 TNKFTNTSGFRANKTQDVLVAQVDFGLRPSIAYTKSKAKDVGIGDVLVNFVGAT 322
 |||||||
 QY 301 YFENKMSYVDYTIINQIDSDNKLGVGSDDTVAAGIYQF 340
 |||||||
 DB 323 YFENKMSYVDYTIINQIDSDNKLGVGSDDTVAAGIYQF 362
 |||||||

RESULT 2
 PHOE_ECOLI STANDARD; PRT: 351 AA.
 AC P02932;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
 GN PHOE OR OMP.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN 11
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=63189086; Pubmed=6341601;
 RA Overbeek N., Bergman H., van Mansfeld F., Lugtenberg B.;
 RT "Complete nucleotide sequence of phoe, the structural gene for the
 RT phosphate limitation inducible outer membrane pore protein of
 RT Escherichia coli K12.";
 RL J. Mol. Biol. 163:513-532(1993).
 RN 12
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RA Takemoto K., Mori H., Murayama N., Katano K., Yano M., Itoh T.,
 RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
 RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
 RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
 RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN 13
 RC SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=9742617; Pubmed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN 14
 RC SEQUENCE FROM N.A.
 RP Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,

RA Davis K., Federpiel N., Hyman R., Kalman S., Komp C., Kurli O.,
 RA Lashkari D., Lew H., Lin D., Namath A., Oelner P., Roberts D.,
 RA Davis R.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN 15
 RP SEQUENCE OF 1-22 FROM N.A.
 RX MEDLINE=84297233; Pubmed=6089111;
 RA Deutch A.H., Rushlow K.E., Smith C.J.;
 RT "Analysis of the Escherichia coli probA locus by DNA and protein
 RT sequencing.";
 RL Nucleic Acids Res. 12:6337-6355(1984).
 RN 16
 RP MUTAGENESIS OF PHE-351.
 RX MEDLINE=91162638; Pubmed=1848301;
 RA Struyve M., Moons M., Tommassen J.;
 RT "Carboxy-terminal phenylalanine is essential for the correct assembly
 RT of a bacterial outer membrane protein.";
 RL J. Mol. Biol. 218:141-148(1991).
 RN 17
 RP X-RAY CRYSTALLOGRAPHY (6.0 ANGSTROMS).
 RX MEDLINE=91172301; Pubmed=1848682;
 RA Jap B.K., Wallan P.J., Gehring K.;
 RT "Structural architecture of an outer-membrane channel as determined
 RT by electron crystallography.";
 RL Nature 350:167-170(1991).
 RN 18
 RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
 RX MEDLINE=92375189; Pubmed=1380671;
 RA Cowan S.W., Schirmer T., Rummel G., Stelert M., Ghosh R.,
 RA Pauprit R.A., Jansonius J.N., Rosenbusch J.P.;
 RT "Crystal structures explain functional properties of two E. coli
 RT porins.";
 RL Nature 358:727-733(1992).
 RN 19
 RP TOPOLOGY.
 RX MEDLINE=93172954; Pubmed=7679770;
 RA Struyve M., Visser J., Adriaanse H., Benz R., Tommassen J.;
 RT "Topology of Phoe porin: the 'eyelet' region.";
 RL Mol. Microbiol. 7:131-140(1993).
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN E. COLI CELLS
 CC ARE GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
 CC SOLUTES.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: V00316; CA23605.1; -
 CC EMBL: D83536; BAA7910.1; -
 CC EMBL: AE000132; AAC73345.1; -
 CC EMBL: U70214; AAB08661.1; -
 CC EMBL: X00786; CA25362.1; -
 CC PIR: A03432; MMECPK.
 CC PDB: 1PHO; 3I-OCT-93.
 CC SWISS-2DPAGE: P02932; COLI.
 CC ECO2DBASE: B037.0; 6TH EDITION.
 CC EcoGene: EG10729; PHOE.
 CC InterPro: IPR001702; -
 CC InterPro: IPR001897; -
 CC Pfam: PF00267; Gram-ve_porins; 1.
 CC PRINTS: PR00182; ECOLIPORIN.
 CC PRINTS: PR00183; ECOLIPORIN.
 CC PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KW Outer membrane; Transmembrane; Porin; Signal; 3D-structure.

```

FT SIGNAL 1 21
FT CHAIN 22 351
FT MUTAGEN 351 351
FT STRAND 25 26
FT STRAND 30 35
FT STRAND 37 45
FT STRAND 52 53
FT STRAND 56 65
FT STRAND 71 82
FT STRAND 87 88
FT STRAND 93 104
FT TURN 105 107
FT STRAND 108 116
FT TURN 119 119
FT HELIX 120 123
FT TURN 124 126
FT TURN 140 141
FT TURN 143 145
FT STRAND 147 157
FT TURN 159 162
FT STRAND 166 173
FT STRAND 176 176
FT STRAND 182 183
FT STRAND 185 185
FT STRAND 188 197
FT TURN 198 201
FT STRAND 202 212
FT HELIX 215 217
FT TURN 218 219
FT STRAND 227 239
FT TURN 240 241
FT STRAND 242 252
FT STRAND 264 274
FT STRAND 280 292
FT STRAND 294 294
FT TURN 295 297
FT STRAND 298 298
FT STRAND 301 313
FT STRAND 318 327
FT TURN 334 335
FT STRAND 342 350
SQ SEQUENCE 351 AA: 38922 MW: 249E2E362C030C9A CRC64:

Query Match 61.4%; Score 1110; DB 1; Length 351;
Best Local Similarity 62.1%; Pred. No. 3.9e-72;
Matches 213; Conservative 38; Mismatches 76; Indels 16; Gaps 5;

```

```

RESULT 3
PHOE_ENTCL
ID PHOE_ENTCL STANDARD; PRT: 350 AA.
AC Q47490;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Enterobacter cloacae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Enterobacter.
OX NCBI_TaxID=550;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87190417; PubMed=3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the phoE genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475(1987).
CC -! FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICIPATE IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -! SUBUNIT: HOMOTRIMER.
CC -! SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -! SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M28296; AAA24809.1; -.
DR InterPro; IPR001702; -.
DR InterPro; IPR001897; -.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLI_PORIN.
DR PRINTS; PR00183; ECOLI_PORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 350
FT SEQUENCE 350 AA: 38511 MW: 101655FEECC3AFAD CRC64:

Query Match 59.3%; Score 1072.5; DB 1; Length 350;
Best Local Similarity 60.3%; Pred. No. 1.8e-69;
Matches 207; Conservative 40; Mismatches 79; Indels 17; Gaps 5;

```

```

Db 254 TPI-----SGGFANKQNFVNAQYDFDGLRSLGVSQKGNQNEIGEDDLTKYIDV 307
OY 298 GATTFYFNKMSSTVYDIINQIDSNKLCVSDDTVANGIYQF 340
Db 308 GATTFYFNKMSAFVDYKINQIDDNKLCVSSDDIYAVGMTYQF 350

RESULT 4
OMPF_SERMA
ID OMPF_SERMA STANDARD: PRT: 374 AA.
AC 033980;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PROTEIN F PRECURSOR (PORIN OMPF).
GN OMPF.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UOC-51;
RX MEDLINE=97419518; PubMed=9274033;
RA Hutsul J.A.M., Morobec E.A.;
RT "Molecular characterization of the Serratia marcescens OmpF porin,
RT and analysis of S. marcescens OmpF and OmpC osmoregulation.";
RL Microbiology 143:2797-2806(1997).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPF/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U81967; AAB69103.1; ALT_INIT.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins. 1.
DR PRINTS: PR00182; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN. 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 374 OUTER MEMBRANE PROTEIN F.
SQ SEQUENCE 374 AA; 41184 MW; BF821D385E441049 CRC64;

Query Match 59.3%; Score 1072; DB 1; Length 374;
Best Local Similarity 60.1%; Pred. No. 2,1e-69;
Matches 215; Conservative 38; Mismatches 81; Indels 24; Gaps 10;
OY 1 AEIYNKQGNKVDLYGKAVGLHYFSKNGENSGYSGDMTVARLFGKGTQINSDLTGYGQ 60
Db 23 AEIYNKQGNKVDLYGKAVGLHYFSKNGENSGYSGDMTVARLFGKGTQINSDLTGYGQ 77

OY 61 WEYNFQGNSESGADQAGTGNKRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPFEGG 120
Db 78 WEYNVQSNHSAOSQTE-GTKTRLGFAGLKFADYGSFDYGRNYGVLYDEGWTMDLPFEGG 136

OY 121 DT-AYSDFEFGRGVAVTAYRNSNPEGLVNDGLNFAVQYLGNERNB--TARSNDGCVGGS 177
Db 137 DTYVYSDMEFTGRNGVATYNNNFGLVLDGLNPLALQYQGNQNDGRDVKRQNGDGMGWS 196

OY 178 ISYEY-EGFGIVAGYAGADRNLQEAOPLGNGKRAEQMATGLKYDANNIYLAANGETRNN 236
Db 197 STYIDGEGVSGAAYASSNRDQDKLRSENGKADAMTVGAKKDANNVYLAAMYAEIRN 256

```

```

OY 237 AFPI-TNKFTNT-----SGFANKTODVLVAQYDFDGLRSLGVSQKGNQNEIGEDDLTKYIDV 286
Db 257 MTFPGGGFTNTGTCATNCGGFASTQNFVTAQYDFDGLRSLGVSQKGNQNEIGEDDLTKYIDV 316

OY 287 G-DVDLVNPFVFGATYYENKMSSTVYDIINQIDSDN--KLGVSSDDTVANGIYQF 340
Db 317 GSDQDLVAVSVGTFYYENKMSSTVYDYKINLDDNDFTKATGATDIDYGVGLYQF 374

RESULT 5
PHOE_CITFR
ID PHOE_CITFR STANDARD: PRT: 351 AA.
AC 001605;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Sierlings G., Ockhuysen C., Hoisra H., Tommassen J.;
RT "Characterization of the Citrobacter freundii phoE gene and
RT development of C. freundii-specific oligonucleotides.";
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDICED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPF/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X68021; CAA48162.1; -.
DR PIR: S25520; S25520.
DR HSP: P02932; IPHO.
DR InterPro: IPR001702; -.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins. 1.
DR PRINTS: PR00182; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN. 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 351 AA; 38807 MW; 4E9C144070381B5A CRC64;

Query Match 59.1%; Score 1070; DB 1; Length 351;
Best Local Similarity 59.2%; Pred. No. 2,7e-69;
Matches 203; Conservative 48; Mismatches 76; Indels 16; Gaps 5;
OY 1 AEIYNKQGNKVDLYGKAVGLHYFSKNGENSGYSGDMTVARLFGKGTQINSDLTGYGQ 60
Db 22 AEIYNKQGNKVDLYGKAVGLHYFSKNGENSGYSGDMTVARLFGKGTQINSDLTGIGR 76

OY 61 WEYNFQGNSESGADQAGTGNKRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPFEGG 120
Db 77 WEAFEGAGKASDSNQ--QKTRLAFAGSKLNLSGFDYGRNGLALYDEAMTVDFPFEFG 134

```

```

Oy 121 D-TAYSDPEVGRGCVATATRSNPFELVDGLNFAVYLGKNEEDTARSRNOCVGSIS 179
Db 135 DSSAQOTNEFTKRASGLATYRNTDFEFGVGLDLTLQYQGRKNDRYDKQKQNGDFGTSVT 194
Oy 180 YEYAG--FGIYVAGGADRTNLCQAPFLGNGKAEQWATGLKYDANNITYLAANYGETRNA 237
Db 195 YDEGSDPFAVSGATNTSDRINQONLQTRGDKAKAEAWATGLKYDANDIYATFSETRNM 254
Oy 238 TPITNKTFTNTSGFANQTQDVLLVAQYQFDFGLRPSITATYSKADVSGICGVDLVNFEV 297
Db 255 TPI-----SGGFANKTQNEFAVYQYQFDFGLRPSLGYLSKGDIGGVNEDLVNVIDV 308
Oy 298 GATYFFNKNMSTVYDIITINQIDSNKTCGVGSDVVAAGIYQF 340
Db 309 GATYFFNKNMSAFYDIKINQIDSDNKLGINNDIVVAGMYQF 351

RESULT 6
PHOE_SALTY STANDARD; PRT; 350 AA.
ID PHOE_SALTY STANDARD; PRT; 350 AA.
AC P30705;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=93083994; PubMed=1280609;
R1 Spleetings G., Elders R., van Lith B., Hofstra H., Tommassen J.;
R7 "Characterization of the Salmonella typhimurium phoE gene and
R7 development of Salmonella-specific DNA probes.";
R7 Gene 122:45-52(1992).
CC -!- FUNCTION: THIS IS ONE OF THE PROTEINS INDICED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORILATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -!- SUBUNIT: HOMOTRIMER.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE OMP/C/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68023; CAA48164.1; -.
DR PIR: S25525; S25525.
DR HSSP: P02932; IPHO.
DR StyGene: SG10291; PHOE.
DR InterPro: IPR001702; -.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLIPEPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1
FT CHAIN 20
SQ SEQUENCE 350 AA; 38762 MW; 90IEE3355AF59877 CRC64;

```

QY	1	AEIYNKDGKNVDILYKKAVALGHLFYSKSGENSENSYGGNGDMTYARLGFSGETQINSDLITGQG	60
Db	21	AEVYNKNGKNRLDYGKVKAMHYMSDYDSK-----DGDQSYVRFEGFGKGETQINDLTGGR	75
QY	61	WEYFQGNNSGADAOQGNKTRLAPAGLKKYADVDSFDGRNVGVYDLAGTMDLPERFG	120
Db	76	WEAFASNAKASDSSQ--QKTRLAPAGLKKLIDGSEFDGRNLGALYDAWMTDMPERFG	133
QY	121	D-TPYSDDFEFGRGVATYRNSNFEGLVDLNFVAOYLKGNERNPTARRNSGDVGSGIS	179
Db	134	DSQAQTDNFMKRRASGLATYRNTDFEGLVDGLDITLQYQGNKEDRDVAKQNGDGGFTSYFS	193
QY	180	YEEYEG--FGIVGAGAADRTMLQEAQPLGNGKKAKQWATGLKYDANNIYLAANYGETRRA	237
Db	194	YDFGSDPEFAVGAATLTSDRTREONLQRRGTGDDKAEAMATGVYDANDIYIAFYSETRM	253
QY	238	TPITNKFNTSGFANKQDVLLVVAQYQDFGRPSIATYTKSAKDVEGIGDVLVNFYEV	297
Db	254	TPV-----SGGFARKTQNFENAVIYQYDFGRPSLGYLASKGRIEYGSSEDLVNIYD	307
QY	298	GATYFFNKNMSTYVDYILNQIDSDNKLGVSGDDYFAVGIVYQF	340
Db	308	GATYFFNKNMSAFYDYKINQDLSDDTLGINDDIYALIGLTYQF	350

RESULT 7
 ID OMPC_ECOLI STANDARD: PRT: 367 AA.
 AC P06966:
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OUTER MEMBRANE PROTEIN C PRECURSOR (PORIN OMPC) (OUTER MEMBRANE PROTEIN 1B).
 DE OMPC OR MECA OR PAR.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=83213433; PubMed=6304064;
 RA Mizuno T., Chou M.-Y., Inouye M.;
 RT "A comparative study on the genes for three porins of the Escherichia coli outer membrane. DNA sequence of the osmoregulated ompc gene.";
 RL J. Biol. Chem. 258:6932-6940(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12:
 RX MEDLINE=97251358; PubMed=9097040;
 RX Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Raza Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y., Sivasubram S., Tagami H., Takeeda J., Takemoto K., Mada C., Yamamoto Y., Horichi T.;
 RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 RN [4]

RP SEQUENCE OF 218-367 FROM N.A.
RC STRAIN-K12 / BHB2500.
RA Richerich P., Lacey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RL Church G.M.;
RN Submitted (OCr-1993) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE OF 1-22 FROM N.A., AND SEQUENCE OF 22-40.
RX MEDLINE-83132326; PubMed-6297988;
RA Mizuno T., Chou M.-Y., Inouye M.;
RT "DNA sequence of the promoter region of the ompC gene and the amino
RT acid sequence of the signal peptide of pro-ompC protein of
RT *Escherichia coli*.";
RL Peps Lett. 151:159-164(1983).
RN [6]
RP SEQUENCE OF 32-57 FROM N.A.
RX MEDLINE-86033642; PubMed-2997131;
RA Nogami T., Mizuno T., Mizushima S.;
RT "Construction of a series of ompC-ompC chimeric genes by in vivo
RT homologous recombination in *Escherichia coli* and characterization of
RT the translational products.";
RL J. Bacteriol. 164:797-801(1985).
RN [7]
RP SEQUENCE OF 22-30.
RX STRAIN-K12 / EMG2;
RC MEDLINE-97443975; PubMed-9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of *Escherichia coli* K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [8]
RP SEQUENCE OF 22-26.
RC STRAIN-K12 / W3110;
RX MEDLINE-98291876; PubMed-9629924;
RA Molloy M.P., Herbert B.R., Walsh B.J., Tyler M.I., Traini M.,
RA Sanchez J.-C., Hochstrasser D.F., Williams K.L., Gooley A.A.;
RT "Extraction of membrane proteins by differential solubilization for
RT separation using two-dimensional gel electrophoresis.";
RL Electrophoresis 19:837-844(1998).
CC -1- FUNCTION: FORM PASSIVE DIFFUSION PORES WHICH ALLOW SMALL MOLECULAR
CC WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K00541; AAA24243.1; -
DR EMBL: AE000310; AAC75275.1; -
DR EMBL: D90850; BAA15998.1; -
DR EMBL: U00008; AAA16412.1; -
DR PIR: A20867; MMECPG.
DR PIR: B25029; B25029.
DR PIR: A18885; A18885.
DR HSSP: P02931; 1GFM.
DR ECODBASE: A035.5; 6TH EDITION.
DR Ecogene; EG10670; ompC.
DR InterPro: IPR001702; -
DR InterPro: IPR001897; -
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLEIPORIN.
DR PRINTS: PR00183; ECOLEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Phage recognition; Signal.
FT SIGNAL 1 21
FT CHAIN 22 367 OUTER MEMBRANE PROTEIN C.
SO SEQUENCE 367 AA; 40368 MW; 6AA9370CC8A1A225 CRC64;

Query Match 58.1%; Score 1051; DB 1; Length 367;
Best Local Similarity 58.9%; Pred. No. 6,4e-68;
Matches 212; Conservative 45; Mismatches 69; Indels 34; Gaps 8;
QY 1 AEIYNKDGKRVLDYKRAKGYHFSKNGENSYGNGDMYARLGFKEGTQINSIDLTGSGQ 60
DB 22 AEYNNKDGKRLDLYKGYDGLHFS-----DNKDYDQITMRGLRKEGTQVTLQITGQ 76
QY 61 WEYNFGNNSEGADQGTNRLAFAGLKYADYSGDFGNNGYVYDALGYTDLMPFEGG 120
DB 77 WEYQIGNSAEN--ENMSYTRVAFAGLKFQDYGSPFGNNGYVYDVTMTDLPFEGG 133
QY 121 DTAYSDDFEYGRYGVATYNSNPFGLVGLNFAVQYLGKNERDT-----A 166
DB 134 DTYSDFNFQORNGEYATYNTDFEGLVGLNFAVQYGRNGNPSGEGFTSGVTNNGRDA 193
QY 167 RRSNGDVGGSISYEYEGFIVGAYGADRTNLOE-AQPGNGKKAKOMATGLKYDANNI 225
DB 194 LRQNGDVGGSITYDIEGFEGFICGAISSKRTDQONTAAYTGNGDRATYTGGLKYDANNI 253
QY 226 YLAANYGETRNATPTNKKFTNTSGFANKTQDVLVAQYDFGLRPSIAYTKSRANDV-E 284
DB 254 YLAQYTGTYNATRV-----GSLGMANKAQNFQAVQYDFGLRPSIAYTQSKGNLGR 308
QY 285 GICDVLVNTFEVQATYENKKNSTIYDIIINOISDNKL-----GVSDDTVAVGIYQF 340
DB 309 GYDEDEILKIVDQATYFENKKNSTIYDVKINLID-DNQETPDAGINTDVIVALGIYQF 367
RESULT 8
ID PROE_KLEPN STANDARD; PRT; 351 AA.
AC P30704;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PROE.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-87190417; PubMed-3032618;
RA van der Ley P., Bekkers A., van Meersbergen J., Tommassen J.;
RT "A comparative study on the pho genes of three enterobacterial
RT species. Implications for structure-function relationships in a
RT pore-forming protein of the outer membrane.";
RL Eur. J. Biochem. 164:469-475(1987).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M28295; AAA25121.1; -
DR HSSP: P02932; IPHO.
DR InterPro: IPR001702; -
DR InterPro: IPR001897; -
DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLIPEPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 351 OUTER MEMBRANE PORE PROTEIN E.
SQ SEQUENCE 351 AA; 38894 MW; 86A5286C12502EC7 CMC64;

Query Match 58.0%; Score 1050; DB 1; Length 351;
Best Local Similarity 58.6%; Pred. No. 7.2e-68;
Matches 201; Conservative 43; Mismatches 83; Indels 16; Gaps 5;

QY 1 AEIYKDKNKVDLYGKAVGLHYFSKNGENSTYGGDMTYARLGKGETQINSDLTGIGQ 60
DB 22 AEVYKKNKNDLYGKIKAMHYFSDDSK-----DGDQTYVYFGIKGETQINEDLTGYGR 76
QY 61 WEYNQGNNSGADQOTGNKTRTLAGLKYADVSGFDYGRNRYVYDALGYTDMPEFGG 120
DB 77 WESEFSGKTESDSSQ--QKTRTLAGVYKLNKYSFSDYGRNRYVYDALGYTDMPEFGG 134
QY 121 D-TAYSDPEFVGVRGVATYRNSNFFGLVDGLNFAVQYLGKNERDTARRNSGDSGVSIS 179
DB 135 DSSAQTDNFMKRRASGLATYRNTDFEGLVDGLDLYOGKNGEAKKQNGDGYGTSLIS 194
QY 180 YEEEG--FGIVGAGAADRTNLQEAQPLGNGKKAQOMATGKLYDANNITYLAANYGETRNA 237
DB 195 YDFGSTDFAVSAAYTSRDTNDQNLARAGSKAMATGLKYDANNITYLAANYSETRKM 254
QY 238 TPITNKFTNTSGFAKKTQDVLVLAQYDFGLRPSIATYKSKAKVEGIDVDLYNREV 297
DB 255 TPT-----SGGFKAQKNFEAVAYQDFGRLPSLGLVYLSGKKEIEGVSIEDLVYIDV 308
QY 298 GATYFFNKMSTYVDYIINQIDSDNKLGVSDDTVAAGIVYQF 340
DB 309 GUTYFFNKNMNAFVYDKINQKSDMKLGINDDIYALGMYTGF 351

RESULT 9
PORI_BPPA2 STANDARD; PRT; 365 AA.
AC 007238;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.
GN LC.
OS Bacteriophage PA-2.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
OX NCBI_TaxID=10738;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86304457; PubMed=3017988;
RA Blassband A.J., Marcotte W.R. Jr., Schnaltman C.A.;
RT "Structure of the lc and mmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage."
RL J. Biol. Chem. 261:12723-12732(1986).
CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES
CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE
CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER
CC THAN 4000-6000 DALTONS).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE
CC (BY SIMILARITY).
CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF
CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN
CC THE LYSOGENIC STATE. IN E. COLI, THE EXPRESSION OF OMPC AND OMPF
CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL: J02580; AAA32301.1; -.
DR PIR: D25647; MMAPR2.
DR HSSP: P02931; IGEN.
DR InterPro: IPR001702; -.
DR InterPro: IPR001897; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLIPEPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORE PROTEIN LC.
FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).
SQ SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CMC64;

Query Match 57.8%; Score 1046; DB 1; Length 365;
Best Local Similarity 59.5%; Pred. No. 1.4e-67;
Matches 210; Conservative 35; Mismatches 84; Indels 24; Gaps 7;

QY 1 AEIYKDKNKVDLYGKAVGLHYFSKNGENSTYGGDMTYARLGKGETQINSDLTGIGQ 60
DB 24 AEIYKKNKNDLYGKIKAMHYFSNDAD-----DGDYTYARLGKGETQINDLDTGIGQ 78
QY 61 WEYNQGNNSGADQOTGNKTRTLAGLKYADVSGFDYGRNRYVYDALGYTDMPEFGG 120
DB 79 WEYEFKGNRAE--SGSSSKDKTHLAPAGLKFQDYGSDYGRNRYVYDALGYTDMPEFGG 137
QY 121 DT-AYSDFEFGVRGVATYRNSNFFGLVDGLNFAVQYLGKNER--DTARRNSGDSGVS 176
DB 138 DTWQTDVFMGRITGFADRYRNDPEGLVDGLNFAVQYLGKNDSDDFNYTEGNGDGGF 197
QY 177 STSYEEGFGIVGAYGAADRNLQE-----AQPLGNGKKAQOMATGKLYDANNITYLAN 230
DB 198 SATYEEEGFGIGATYAKDSDRTQYNAKRVLPYEVASKNMVAAGKLYDANNITYLATY 257
QY 231 YGETRNATPTNKFTNTSGFANKTQDVLVLAQYDFGLRPSIATYKSKAKVEGIDVD 290
DB 258 YSETQNMVTFADHFV-----ANKAQNFENVAQYDFGLRPSVATLQSKGDLGWGDD 312
QY 291 LVNFEVGATYFFNKMSTYVDYIINQIDSDN--KLGVSDDTVAAGIVYQF 340
DB 313 LVKYVDGATYFFNKMSTYVDYKINLKDNDFTKALGVSTDIYAVAGIVYQF 365

RESULT 10
PHOE_SALTI
ID PHOE_SALTI STANDARD; PRT; 350 AA.
AC 056119;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PORE PROTEIN E PRECURSOR.
GN PHOE.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IMS-1;
RA Torres A., Puente J.L., Calva E.;
RL Submitted (Aug-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.

```

CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X74595; CAA52672.1; -
DR HSP: P02932; 1PHO.
DR InterPro: IPR001702; -
DR pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
DR Outer membrane; Transmembrane; porin; Signal.
KW SIGNAL
FT CHAIN 1 20 BY SIMILARITY.
FT SEQUENCE 350 AA; 38639 MW; 4136D5E4F0CC3A27 CRC64;
SQ
-----
Query Match 57.7%; Score 1044; DB 1; Length 350;
Best Local Similarity 57.4%; Pred. No. 1.9e-67;
Matches 197; Conservative 47; Mismatches 83; Indels 16; Gaps 5;
OY 1 AEIYNKDKKKVDLYGKAVGLHYTSKNGENSYGNGDMYARLGFGEIYDNLGYGQ 60
DB 21 AEIYNKNGKRLDYGVKKAMHYSDDSK-----DDQSYVRKFGKQIYDNLGYGGR 75
OY 61 WEYFNQGNSEGAQDQGNKTRAFAGLKYADYGSFQYGNVYDALCYTDLPEFGG 120
DB 76 WEHGGKNAESSSSQ--QKTRAFAGLKLKIDGSPDYGNLCAILDYEMTMMFPEFGG 133
OY 121 D-TAYSDDFVGVGVAVYRNSNFGVLDGLNFAVOYLGRNDRYARNSGDSGVSIS 179
DB 134 DSSAQDNFMTKRAAGLATYRNTDFGIYDGLDTLOYGKNEDRDYKQNGDGFYSVS 193
OY 180 YEEG--FGIVGAYGADPRNLQEAQPLNGKKAQEAQAGLTKDANNIYLAANYGERNA 237
DB 194 YDGGSDFAVSGAYLTSDRREQNLQRRGTGDAEGMAGGVKKYDADITYATYSETRNM 253
OY 238 TPTNFTTSGFANKTDVLAQYQDFGLRPSIATYTSKAKDEGIGDVLVNYFEV 297
DB 254 TPT-----SGFRANKTQNEAVIYQYDFGLRPSIGLYLSKGRKIDEGVSEDLVNTIDV 307
OY 298 GATYFNKNNSTYVDYIINOIDSNKLGVGSDDTVAVGIVYQF 340
DB 308 GATYFNKNNSAFVDYKINQDLSNDTLGINDDDIVAIGLYTQF 350
-----
RESULT 11
NMPC_ECOLI STANDARD; PRT; 365 AA.
AC P21420; P77189;
DT 01-MAY-1991 (Rel. 18, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE OUTER MEMBRANE PORIN PROTEIN NMPC PRECURSOR.
GN NMPC OR PHMA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MUTANT CS384;
RX MEDLINE=86304457; PubMed=3017988;
RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;

```

```

RT "Structure of the lc and nmpc outer membrane porin protein genes of
RT lambdaoid bacteriophage.";
RT J. Biol. Chem. 261:12723-12732(1986).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:12453-12474(1997).
RL [3]
RP SEQUENCE FROM N.A.
RC STRAIN=JL5502;
RX MEDLINE=94335635; PubMed=8057841;
RA Coll J.L., Heyde M., Portallier R.;
RT "Expression of the nmpc gene of Escherichia coli K-12 is modulated by
RT external pH. Identification of cis-acting regulatory sequences
RT involved in this regulation.";
RT Mol. Microbiol. 12:83-93(1994).
RL [5]
RP SEQUENCE OF 347-365 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=96196428; PubMed=8648624;
RA Mahdi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;
RT "Holliday junction resolvases encoded by homologous rusa genes in
RT Escherichia coli K-12 and phage 82.";
RT J. Mol. Biol. 257:561-573(1996).
RL [1]
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, OUTER MEMBRANE.
CC MISCELLANEOUS: IN WILD-TYPE STRAINS OF E. COLI K12, THE NMPC OPEN
CC READING FRAME IS INTERRUPTED BY AN ISS INSERTION AND GENERATES A
CC HYBRID OPEN READING FRAME THAT IS NOT EXPRESSED. HOWEVER, IN
CC MUTANT STRAIN CS348, THE ISS ELEMENT HAS BEEN DELETED AND NMPC IS
CC EXPRESSED.
CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M13457; AAA23728.1; ALT_SEQ.
DR EMBL: AE000160; AAC73654.1; ALT_SEQ.
DR EMBL: U82598; AAB40749.1; ALT_SEQ.
DR EMBL: Z35442; CAA84594.1; -
DR EMBL: X92587; CAA63325.1; -
DR PIR: A25647; MMECNC.
DR HSP: P02931; 1GFN.
DR Ecocore: EGI0659; nmpc.
DR InterPro: IPR001702; -
DR pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PRINTS: PR00183; ECOLIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Transmembrane; porin; Signal.
FT SIGNAL 1 23
FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN NMPC.
FT CONFLICT 326 326 K -> N (IN REF. 2 AND 3).
SQ SEQUENCE 365 AA; 40316 MW; 6E512BD4847FBAF8 CRC64;

```


[3]
 RN SEQUENCE FROM N.A., SEQUENCE OF 22-33, AND CHARACTERIZATION.
 RP MEDLINE-98317278; PubMed-9642192;
 RA Prilipov A., Pale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
 RT Identification and characterization of two quiescent porin genes,
 RL ompc and ompn, in Escherichia coli BE.";
 RL J. Bacteriol. 180:3388-3392(1998).
 CC -1- FUNCTION: HOMOTRIMER.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AE000234; AAC74459.1; -
 CC EMBL: D90775; BAA14981.1; -
 CC EMBL: D90776; BAA14986.1; -
 CC DR Ecogene: EG13375; ompn.
 CC DR InterPro: IPR001897; -
 CC DR InterPro: IPR001897; -
 CC DR Pfam: PF00267; Gram-ve_porins; 1.
 CC DR PRINTS: PR00182; ECOLNEIPORIN.
 CC DR PRINTS: PR00183; ECOLIPORIN.
 CC Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 377
 FT SEQUENCE 377 AA; 41220 MW; 4420DZCB9BC10F95 CRC64;

Query Match 57.2%; Score 1035; DB 1; Length 377;
 Best Local Similarity 56.9%; Pred. No. 9e-67; Indels 38; Gaps 9;
 Matches 209; Conservative 42; Mismatches 78;

QY 1 AEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 60
 DB AEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 76
 QY 61 WEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 120
 DB WEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 134
 QY 121 DT--AYSDPEFVGRGVATYRNSNFFGLVDGLNFAVOYLGNKNE-----RPTAR 167
 DB DT--AYSDPEFVGRGVATYRNSNFFGLVDGLNFAVOYLGNKNE-----RPTAR 193
 QY 168 RNSGDEGVGSISLEY-EGFQIVGAYGAADRTNLQEAOPLGNGKKAQEWATGLKYDANNIY 226
 DB HEMGDGGLSTYTDLGKGFAGAAIYSSORTNQVNHNTAAGCKADKAWTGLKYDANNIY 253
 QY 227 LAANGGTRATATITNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDE-- 284
 DB LAANGGTRATATITNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDE-- 310
 QY 254 LATMYSSTRMTPFGD---SDVAVANKTQNFETVTAQYQDFGLRPAVSFLMSKGRDLHAA 310
 DB LATMYSSTRMTPFGD---SDVAVANKTQNFETVTAQYQDFGLRPAVSFLMSKGRDLHAA 333
 QY 285 -----GIGDVLVNFVEFGATYFENKKNSTVVDYIINDSDNKL-----GVGSDDTVA 333
 DB -----GIGDVLVNFVEFGATYFENKKNSTVVDYIINDSDNKL-----GVGSDDTVA 370
 QY 334 VGIYQF 340
 DB VGIYQF 377

RESULT 14
 ID OMPF_SALTI STANDARD; PRT; 363 AA.
 AC 056113;
 DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE OUTER MEMBRANE PROTEIN F. PRECURSOR (PORIN OMPc) (OUTER MEMBRANE
 DE PROTEIN S3).
 GN OMPF OR OMP53.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN SEQUENCE FROM N.A.
 RC SPRAIN-IMSS-1;
 RA Fernandez-Mora M., Calva E.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: OMPF IS A PORIN THAT PASSIVE DIFFUSION PORES WHICH ALLOW
 CC SMALL MOLECULAR WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER
 CC MEMBRANE. IT IS ALSO RECEPTOR FOR THE BACTERIOPHAGE T2 (BY
 CC SIMILARITY).
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE OMPc/PHOE FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X8957; CAA61905.1; -
 CC DR InterPro: IPR001897; -
 CC DR InterPro: IPR001897; -
 CC DR Pfam: PF00267; Gram-ve_porins; 1.
 CC DR PRINTS: PR00182; ECOLNEIPORIN.
 CC DR PRINTS: PR00183; ECOLIPORIN.
 CC DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 CC Outer membrane; Transmembrane; Porin; Phage recognition; Signal.
 FT SIGNAL 23 363
 FT CHAIN 23 363
 FT SEQUENCE 363 AA; 40048 MW; F6069B34E9516859 CRC64;

Query Match 57.0%; Score 1031.5; DB 1; Length 363;
 Best Local Similarity 57.6%; Pred. No. 1.5e-66;
 Matches 200; Conservative 54; Mismatches 80; Indels 13; Gaps 8;

QY 1 AEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 60
 DB AEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 78
 QY 61 WEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 120
 DB WEYNNKDKNVDLYGKAVGLHPSKNGENSGYSGNDMTYARLGFKGTQINSDLTGEGQ 137
 QY 121 DT--AYSDPEFVGRGVATYRNSNFFGLVDGLNFAVOYLGNKNEPTARRS-NGDGVG 176
 DB DT--AYSDPEFVGRGVATYRNSNFFGLVDGLNFAVOYLGNKNEPTARRS-NGDGVG 197
 QY 177 SISIEYFPGIVGAYGAADRTNLQEAOPLGNGKKAQEWATGLKYDANNIYLAANGGTRN 236
 DB SISIEYFPGIVGAYGAADRTNLQEAOPLGNGKKAQEWATGLKYDANNIYLAANGGTRN 256
 QY 198 TMAIEFPGFVYTAIYSSKRTNDQDND--GNGDRAESMAVAGAKDANNVLAAYATETRN 256
 DB TMAIEFPGFVYTAIYSSKRTNDQDND--GNGDRAESMAVAGAKDANNVLAAYATETRN 295
 QY 237 ATPITNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDEGI-GVDLVNPF 295
 DB ATPITNKFTNTSGFANKTDVLLVAQYQDFGLRPSIATYKSKAKDEGI-GVDLVNPF 316
 QY 257 MSIVENTYTDVEANKTQNLVLAQYQDFGLRPAISYQSKGKQNLGAGGSDAKIYI 316
 DB MSIVENTYTDVEANKTQNLVLAQYQDFGLRPAISYQSKGKQNLGAGGSDAKIYI 340
 QY 296 EVGATYFENKKNSTVVDYIINDSDNKLK--VGSDDTVAVGIYQF 340
 DB EVGATYFENKKNSTVVDYIINDSDNKLK--VGSDDTVAVGIYQF 363
 QY 317 QAGATYFENKKNSTVVDYIINDSDNKLK--VGSDDTVAVGIYQF 363

RESULT 15

OMP_C_KLEPN STANDARD: PRT: 363 AA.
 ID OMP_C_KLEPN
 AC Q48473;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE OUTER MEMBRANE PROTEIN C PRECURSOR (PORIN OMP_C) (PORIN OMPK36).
 GN OMP_C OR OMPK36.
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Klebsiella.
 CX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=C3;
 RA MEDLINE=95172739; PubMed=7868262;
 RA Albert S., Rodriguez-Quijones F., Schirmer T., Rummel G., Tomas J.M.,
 RA Rosenbusch J.P., Benedi V.J.;
 RT "A porin from Klebsiella pneumoniae: sequence homology, three-
 RT dimensional model, and complement binding.";
 RL Infect. Immun. 63:903-910(1995).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
 RX MEDLINE=99216538; PubMed=10196126;
 RX Dutzler R., Rummel G., Albert S., Hernandez-Alles S., Phale P.S.,
 RA Rosenbusch J.P., Benedi V.J., Schirmer T.;
 RT "Crystal structure and functional characterization of OmpK36, the
 RT osmoporin of Klebsiella pneumoniae.";
 RL Structure 7:425-434(1999).
 CC -I- FUNCTION: FORM PASSIVE DIFFUSION PORES WHICH ALLOW SMALL MOLECULAR
 CC WEIGHT HYDROPHILIC MATERIALS ACROSS THE OUTER MEMBRANE. IN
 CC K. PNEUMONIAE IT HAS BEEN SHOWN TO BIND C1Q COMPONENT AND ACTIVATE
 CC THE CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM.
 CC -I- SUBUNIT: HOMOTRIMER.
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -I- SIMILARITY: BELONGS TO THE OMP_C/PHO FAMILY OF PORINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 233506; CAA83913.1; .
 DR DR PDB: 1OSM: 26-JUL-99.
 DR InterPro: IPR001702; .
 DR InterPro: IPR001897; .
 DR Pfam: PF00267; Gram-ve_porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PRINTS: PR00183; ECOLIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Transmembrane; Porin; signal; 3D-structure.
 FT SIGNAL 1 21
 FT CHAIN 1 363 OUTER MEMBRANE PROTEIN C.
 FT SEQUENCE 363 AA: 39663 MW: 99DE45546F01F116C CRC64;
 Query Match 56 9%; Score 1030; DB 1; Length 363;
 Best Local Similarity 59.0%; Pred. No. 2e-66; Indels 26; Gaps 7;
 Matches 209; Conservative 37; Mismatches 82;
 OY 1 AETYNKDGKVKVDLYGKAVKALHYFSKNGENSYGNGDMTYARLGFKGETQINSIDLTYGQ 60
 DB 22 AETYNKDGKVKLDLYGKIDLHYFS-----DDKDYDGDQTYMRGKVGKGTQINDQLTYGQ 76
 OY 61 WEXNFGNNSSEGADAGTGNKTRLAFAGLKRYADVGSFIDYGRNKGVVYDALGYTDMLPFGG 120
 DB 77 WEXNFGNNTSSDQAM--TRLAFAGLKFGDAGSFDYGRNKGVVYDVTSMIDVLPFGG 134
 OY 121 DTAISDFEVRGKGVATYRNSNFEGLVDGLNPAVOYLKKN-----ERDT-----ARRSN 170
 DB 171 GDGCGSGISYE-YEGFQIVGAYGAADRTNLPDAPFLGKKAEDQMATLTKTDANNIYLA 229
 DB 195 GDGFGTSYTYDIFDQISAGFAVANSKRRTDDQNQLLGEGDHAETTYTGLKTDANNIYLAT 254
 OY 230 NCGERNATPTITNKRTNTSGPANKTQDYLVAQYQFDEGLRPSIATYSKAKDVEGIDV 289
 DB 255 QTYQYNNR-----AGSLGFPAKQNEFAVAQYQFDFGLRPSVAIYLSKGRDLNGYGDQ 309
 OY 290 DLVNYFEVGAATYFENKNNSTYVDYIINOIDSDN---KLGVSDDTVAVGIVYQF 340
 DB 310 DILKVDYDGAATYFENKNNSTYVDYKINLLDQNSFTRSAIGISTDDVVALGLVYQF 363

DB 135 DTYGSDNFLQSRANQVATYRNSDFRGLVDGLNPAALQYQGNKGSYSGEGATNNGRCALKON 194
 OY 171 GDGCGSGISYE-YEGFQIVGAYGAADRTNLPDAPFLGKKAEDQMATLTKTDANNIYLA 229
 DB 195 GDGFGTSYTYDIFDQISAGFAVANSKRRTDDQNQLLGEGDHAETTYTGLKTDANNIYLAT 254
 OY 230 NCGERNATPTITNKRTNTSGPANKTQDYLVAQYQFDEGLRPSIATYSKAKDVEGIDV 289
 DB 255 QTYQYNNR-----AGSLGFPAKQNEFAVAQYQFDFGLRPSVAIYLSKGRDLNGYGDQ 309
 OY 290 DLVNYFEVGAATYFENKNNSTYVDYIINOIDSDN---KLGVSDDTVAVGIVYQF 340
 DB 310 DILKVDYDGAATYFENKNNSTYVDYKINLLDQNSFTRSAIGISTDDVVALGLVYQF 363

Search completed: July 3, 2001, 15:10:51
 Job time: 751 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:41 ; Search time 95.41 Seconds

(without alignments)
472.865 Million cell updates/sec

Title: US-09-490-291-9
Perfect score: 1809
Sequence: 1 AEIYNKDGKNDLXGKAVGL.....NKLGVSGDDPVAVGVYQFA 341

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 133305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SPTREMBL.16.*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organellar:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-unclassified:*
13: sp-vertebrate:*
14: sp-virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1044.5	57.7	374	2	087754 klebsiella
2	1036	57.3	374	2	085030
3	1029.5	56.9	364	2	09K597 escherichia
4	1029.5	56.9	378	2	09K386 salmonella
5	1028.5	56.9	366	2	09RH85 escherichia
6	1022.5	56.5	333	2	09RAM3 klebsiella
7	1022.5	56.5	333	2	09RAM2 klebsiella
8	1022.5	56.5	363	2	09K3E7 salmonella
9	1021.5	56.4	333	2	09RR59 calymmatoba
10	1019.5	56.4	333	2	09S613 klebsiella
11	1017.5	56.2	360	2	09FR89 salmonella
12	934.5	51.7	359	2	087753 klebsiella
13	895	45.1	315	2	09EXH8 klebsiella
14	770	42.6	315	2	09EC71 yersinia pe
15	428.5	23.7	130	2	052641 yersinia ag
16	310	17.1	111	2	P94857 klebsiella
17	310	17.1	111	2	P94858 klebsiella
18	310	17.1	111	2	P94860 klebsiella
19	309	17.1	111	2	P94635 calymmatoba

20	260.5	14.4	351	2	09RNA1	09nal aeromonas h
21	233.5	12.9	96	2	P77519	P77519 escherichia
22	224	12.4	331	2	09KRU2	09KRU2 vibrio chol
23	197.5	10.9	330	2	P94743	P94743 elkenella c
24	170.5	9.4	69	2	P76871	P76871 escherichia
25	167.5	9.3	329	2	054350	054350 pasteurella
26	162.5	9.0	333	2	09R2D4	09R2D4 pasteurella
27	162.5	9.0	333	2	054339	054339 pasteurella
28	160.5	8.9	329	2	054345	054345 pasteurella
29	159	8.8	323	2	09X586	09X586 neisseria m
30	158	8.7	323	2	066067	066067 neisseria m
31	157	8.7	361	2	045106	045106 burkholderi
32	155.5	8.6	50	2	09EC72	09EC72 yersinia pe
33	155.5	8.6	322	2	051273	051273 neisseria m
34	155.5	8.6	323	2	054343	054343 pasteurella
35	154	8.5	332	2	P72072	P72072 neisseria f
36	153.5	8.5	333	2	054344	054344 pasteurella
37	153	8.5	339	2	051947	051947 neisseria g
38	152.5	8.4	332	2	P72113	P72113 neisseria s
39	152.5	8.4	343	2	054340	054340 pasteurella
40	152	8.4	323	2	070078	070078 neisseria m
41	152	8.4	328	2	054347	054347 pasteurella
42	152	8.4	328	2	054349	054349 pasteurella
43	151.5	8.4	313	2	051266	051266 neisseria m
44	151.5	8.4	322	2	066062	066062 neisseria m
45	151	8.3	327	2	09X587	09X587 neisseria m

ALIGNMENTS

```

RESULT 1
ID 087754 PRELIMINARY; PRT; 374 AA.
AC 087754;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OMPK37 PORIN PRECURSOR.
GN OMPK37.
OS Klebsiella pneumoniae.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD8;
RA Domenech-Sanchez A., Hernandez-Alles S., Martinez-Martinez L.,
RA Benedi V.J., Alberti S.;
RT Identification and characterization of a novel porin of Klebsiella
RT pneumoniae: its role in beta-lactam antibiotics resistance." ;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: A011502; CA009666.1; -
DR HSSP: Q48473; 10SM.
DR InterPro: IPR000408; -
DR InterPro: IPR001702; -
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLEIPORIN.
DR PROSITE: PS00626; RCL_2; UNKNOWN_1.
KW SIGNAL.
FT SIGNAL. 1 21 POTENTIAL.
FT CHAIN 22 374 OMPK37 PORIN.
SQ SEQUENCE 374 AA; 41612 MW; B66F0D74A1A169B CRC64;

```

Query Match 57.7%; Score 1044.5; DB 2; Length 374;
Best Local Similarity 58.5%; Pred. No. 9.3e-69;
Matches 214; Conservative 37; Mismatches 76; Indels 39; Gaps 9;

QY 1 AEIYNKDGKNDLXGKAVGLHYFSKNGENSYSGNGDMFYARLGFKGQINSDLTGYGO 60
DB 22 AEIYNKDGKNDLXGKAVGLHYFSKNGENSYSGNGDMFYARLGFKGQINSDLTGYGO 76

```
QY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNGVYDALGYTDMLPFEFG 120
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 77 WEYNVQANNETSSDQAM--TRLAFAGIKYDGSFDYGRNGVLYLVEGWTDLPEFG 134
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 121 DT-AVSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYLKNERDTA----- 166
| : : : : | | | | | : : : : : | : : : : | : : : : | : : : : |
DB 135 DSTYADNFMAGRANGVATYRNSDFEGLVEGLNFALOYOGKNEGANNODINVCNNSSD 194
| : : : : | | | | | : : : : : | : : : : | : : : : | : : : : |
QY 167 ---RRSNGDVGGSISYEYEGFI--VGAYGADRTNLQEAQPLGSKKAEQWATGLKYD 221
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 195 SDVDFDNDGDFGLSTYDF--CMGISAAAYTSSDRINDQMTQTNARGDKAEMWTAGLKYD 253
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 222 ANNTYLAANGETRNATPITKFTNTSGFANKTQDVLVAQYDFGLRPSIATTKSKAK 281
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 254 ANDYLATMYSERNNMTPYGN----DGVANKTQNFVYTAQYDFGLRPSIATYLSQSKG 308
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 282 DVEGIG---DVDLVNFEVGATYFENKMSYVDYIINOIDSNKL-----GVGSDDTAV 334
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 309 DLYNNGRADDLVKYADVGATYFENRMSYVDYKINLDSGDKFYEDNGISTDNIVAL 368
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 335 GIVYQF 340
| : : : : |
DB 369 GLVYQF 374
```

```
RESULT 2
085030 PRELIMINARY; PRT; 377 AA.
AC 085030;
DB 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PORIN OMPN.
GN OMPN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BE (BL21);
RX MEDLIND-98317278; PubMed=9642192;
RA PiliDPOV A., Phale P.S., Koebnik R., Widmer C., Rosenbusch J.P.;
RT "Identification and characterization of two quiescent porin genes,
nmpC and ompN, in Escherichia coli BE.";
RL J. Bacteriol. 180:3388-3392(1998).
DR EMBL; AF035618; AAC38644.1; -.
HSSP; P02931; IGFN.
DR InterPro; IPR000408; -.
DR InterPro; IPR001702; -.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNIPORIN.
DR PROSITE; PS00626; RCCL_2; UNKNOWN_1.
SQ SEQUENCE 377 AA; 41236 MW; 7E4D681A2BC10F8C CRC64;
```

```
Query Match 57.3%; Score 1036; DB 2; Length 377;
Best Local Similarity 57.2%; Pred. No. 3.9e-68;
Matches 210; Conservative 41; Mismatches 78; Indels 38; Gaps 9;
QY 1 AEIYNKDNKYDLGKAVGLHYFSKNGENSYGNGDMTYARLGFKGFTQINSDLTGYGQ 60
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 22 AEIYNKDKGKLDLYGKVGGLHYFSNDNAK-----DGDOSTALGKRGFTQINDOLTGIGQ 76
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNGVYDALGYTDMLPFEFG 120
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 77 WEYNQANNETSSDQAM--TRLAFAGLKFAVYGSFDYGRNGVYDIEGWTDLPEFG 134
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 121 DT-AVSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYIGKNE-----RDYAR 167
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 135 DSTYADNFMAGRANGVATYRNSDFEGLVEGLNFAVOYOGKNEGASNGOEGTNGRD--VR 193
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 168 RNSGDSVGSISYER--EGGTYGAYGADRTNLQEAQPLGSKKAEQWATGLKYDANNY 226
```

```
DB 194 HENDGGLSTTYDLGGSFAGAAVYTSSDRINDQVNHNTAAGGKADAWTAGLKYDANNY 253
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 227 LAANYGETRNATPITKFTNTSGFANKTQDVLVAQYDFGLRPSIATYKSKAKVE-- 284
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 254 LATMYSETRNNMTPFGD---SDYAVANKTONFEVYTAQYDFGLRPAVSFLMSKGRDHA 310
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 285 -----GIDVDLVNFEVGATYFENKMSYVDYIINOIDSNKL-----GVGSDDTVA 333
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 311 GGADNPAGVDKDLVKYADVGATYFENKMSYVDYKINLDEDSFTYANGISTDDIVA 370
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 334 VGIYQF 340
| : : : : |
DB 371 GLVYQF 377
```

```
RESULT 3
09K597 PRELIMINARY; PRT; 364 AA.
AC 09K597;
DB 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PORIN C PRECURSOR.
GN OMPC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC low A.S.; Mackenzie F.M., Gould I.M., Booth I.R.;
RA "Parallel evolution of multi-resistant bacteria in a patient with
recurrent septicemia: unique data that support the presence of
separate protected environments.";
RT submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
(CY SIMILARITY).
DR EMBL; AJ295721; CAC01403.1; -.
DR InterPro; IPR000504; -.
DR InterPro; IPR001702; -.
DR Pfam; PF00267; Gram-ve_porins; 1.
DR PRINTS; PR00182; ECOLNIPORIN.
DR PROSITE; PS00576; GRAM_NEG_PORIN; 1.
DR PROSITE; PS00030; RNP_1; UNKNOWN_1.
KW Outer membrane; porin; Signal; Transmembrane.
FT SIGNAL 1
FT POTENTIAL.
SQ SEQUENCE 364 AA; 40312 MW; 9B583F2C11344E31 CRC64;
```

```
Query Match 56.9%; Score 1029.5; DB 2; Length 364;
Best Local Similarity 58.3%; Pred. No. 1.1e-67;
Matches 209; Conservative 40; Mismatches 77; Indels 31; Gaps 8;
QY 1 AEIYNKDNKYDLGKAVGLHYFSKNGENSYGNGDMTYARLGFKGFTQINSDLTGYGQ 60
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 22 AEIYNKDKGKLDLYGKVGGLHYFSNDNAK-----DGDYTNKRLGKRGFTQYDOLTGIGQ 76
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 61 WEYNFGNNSGADAOQTGNKTRLAFLAKYADVGSFDYGRNGVYDALGYTDMLPFEFG 120
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 77 WEYQIQGNEPE---SDNSSWTRVAFAGLFDVGSFDYGRNGVYDVTSWTDVLPFEFG 133
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 121 DTAYSDDFVGRGVGATYRNSNFEGLVDGLNFAVOYIGK-----NERDYARS 169
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
DB 134 DTYDSDNFMQDGNFAVYRNTDFGLVDGLFAVOYOGKNGSAHEGKTTNGRDVFEQ 193
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 170 NGDVGGSISYEYEGFIYGAAYGADRT--NLQEAQPLGSKKAEQWATGLKYDANNIYLA 228
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
DB 194 NGDVGGSITTYRBEFGICAANVSSKRTMDQNTGLIGGDAEYITGGLKADANNIYLA 253
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
QY 229 ANYGETRNATPITKFTNTSGFANKTQDVLVAQYDFGLRPSIATYKSKAKV--EGIG 287
| : : : : | | | | : : : : : | : : : : | : : : : | : : : : |
```

Db 254 AAYTQYATNATRV-----GSLGWANKAQNFEAVAYQDFGLRPSLAYLQSKGNKLGRCYD 308

OY 288 DVDLVNFEVQATYFNKKNMSTYVDYIINQIDSDNKL----GVGSDPTVAVGIVYQF 340

Db 309 DEBLKAYDVQATYFNKKNMSTYVDYKINLID-DNRFTRDAGINTDIDVALGLVYQF 364

RESULT 4

OY 09K3E6 PRELIMINARY; PRT; 378 AA.

AC 09K3E6;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN C PRECURSOR.

GN OMP.

OS Salmonella enterica subsp. enterica serovar Minnesota.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

NCBI_TaxID=70803;

OX NCBI_TaxID=70803;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=SF111;

RA Zimmermann H., Maassenaar T.M., Laubenthaler-Preusse H., Petry F.,

RA Loos M.

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE

DR EMBL: Y15844; CAB96613.1; -

DR Interpro: IPR000504; -

DR Interpro: IPR001702; -

DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLNEIPORIN.

DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.

DR PROSITE: PS00030; RNP_1; UNKNOWN_1.

KM Outer membrane; Porin; Signal; Transmembrane.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 378 OUTER MEMBRANE PROTEIN C.

FT SEQUENCE 378 AA; 41269 MW; 254524B9EC3849C CRC64;

Query Match 56.9%; Score 1029.5; DB 2; Length 378;

Best Local Similarity 57.0%; Pred. No. 1.2e-67; Mismatches 78; Indels 33; Gaps 9;

Matches 208; Conservative 46; Mismatches 78; Indels 33; Gaps 9;

OY 1 AEYNNKDNVVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

Db 22 AEYNNKDNKLDLFGKYDGLHYFS-----DDKSGDQOTYMRIGFKETQVNOQLTGYGQ 76

OY 61 WEYNNKDNSEGAADQGNKTRLAFLAGLKYADVGSFDYGRNYGVYDALGYTDLPEFGG 120

Db 77 WEYQIOGNQEGNSN---DSWTRVAFAGLKFADAGSFDYGRNYGVYDVTSMVDLPEFGG 133

OY 121 DTAVSDPEFVGRGVATYRNSNFEGLVDGLNFAVOYLGRK-----ERDTRAR---RSGND 172

Db 134 DTGAGNFMQORNGRNGFYATYRNTDFFGLVDGLNFAVOYLGRKSGVSGNDGRSLNONGD 193

OY 173 DVGSGSISYEY-EGFGIAGVAY----GAADRTNLQEAOPLGKKAQEQATGKTDANNIYL 227

Db 194 GYVGSLTYIYEGFGIAGVAYSSSKRTDDQNS-PLYINGDRAEYTYTGKTKDANNIYLA 252

OY 228 AANYGETRATPTI-----TNKFTNTSGFANKTQVLLVAYQDFGLRPSLAYLQSKGNK 283

Db 254 AAYGSGTYNATRTGTSNGNKNKSDSYGFANKAQNFEAVAYQDFGLRPSLAYLQSKGDI 313

OY 284 E-----GIGVDLVNFEVQATYFNKKNMSTYVDYIINQIDSDN---KLVGSDPTVAVG 335

Db 314 SNGYASYSYGQDILVKYDVQATYFNKKNMSTYVDYKINLID-DNRFTRDAGINTDIDVALG 373

OY 336 IVYQF 340

Db 374 LVYQF 378

RESULT 5

OY 09RH85 PRELIMINARY; PRT; 366 AA.

AC 09RH85;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE OUTER MEMBRANE PROTEIN OMP.

GN OMP.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

NCBI_TaxID=562;

OX NCBI_TaxID=562;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7;

RA Yu S.L., Syu W.J.;

RT "Altered outer membrane protein OmpC in hemorrhagic Escherichia coli

RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE

DR EMBL: AF057355; AF21761.1; -

DR HSSP: Q48473; IOSM.

DR Interpro: IPR001702; -

DR Pfam: PF00267; Gram-ve_porins; 1.

DR PRINTS: PR00182; ECOLNEIPORIN.

DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.

KM Outer membrane; Porin; Signal; Transmembrane.

FT SEQUENCE 366 AA; 40499 MW; 6A4EAD1652565C00 CRC64;

Query Match 56.9%; Score 1028.5; DB 2; Length 366;

Best Local Similarity 58.6%; Pred. No. 1.3e-67; Mismatches 71; Indels 35; Gaps 9;

Matches 211; Conservative 43; Mismatches 71; Indels 35; Gaps 9;

OY 1 AEYNNKDNVVDLYGKAVGLHYFSKNGENSYGNGDMTYARLFGKGTQINSDLTGYGQ 60

Db 22 AEYNNKDNKLDLFGKYDGLHYFS-----DDKSGDQOTYMRIGFKETQVNOQLTGYGQ 76

OY 61 WEYNNKDNSEGAADQGNKTRLAFLAGLKYADVGSFDYGRNYGVYDALGYTDLPEFGG 120

Db 77 WEYQIOGNQSAEN---ENNSWTRVAFAGLKFQDVGSFDYGRNYGVYDVTSMVDLPEFGG 133

OY 121 DTAVSDPEFVGRGVATYRNSNFEGLVDGLNFAVOYLGRK-----NERDTRARSSNG 171

Db 134 DTGSDNFMQORNGRNGFYATYRNTDFFGLVDGLNFAVOYLGRKSGVSGNDGRSLNONG 193

OY 172 DVGSGSISYEYEGFGIAGVAYGAADRTNLQEAOPL-GNGKKAQEQATGKTDANNIYLA 229

Db 194 DVGSGSLTYIYEGFGIAGVAYSSSKRTDDQNS-PLYINGDRAEYTYTGKTKDANNIYLA 252

OY 230 NYGETRATPTIYNKFTNTSGFANKTQVLLVAYQDFGLRPSLAYLQSKGNKGVINGR 287

Db 253 QYGTQYATNATRV-----GSLGWANKAQNFEAVAYQDFGLRPSLAYLQSKGNKGVINGR 307

OY 288 ---DVDLVNFEVQATYFNKKNMSTYVDYIINQIDSDNKL----GVGSDPTVAVGIVYQF 340

Db 308 NYDEDLTKYDVQATYFNKKNMSTYVDYKINLID-DNRFTRDAGINTDIDVALGLVYQF 366

RESULT 6

OY 09RAM3 PRELIMINARY; PRT; 333 AA.

AC 09RAM3;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

DE PHOSPHATE PORIN (FRAGMENT).

GN PHOE.

OS Klebsiella pneumoniae (subsp. ozaenae);
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxId=574;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1436;
 RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
 RT "Phylogenetic analysis of Calymmatobacterium granulomatis.";
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: HOMOTIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 (BY SIMILARITY).
 CC EMBL: AF009229; AAD21517.1; -.
 DR HSSP: P02932; 1PHO.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Porin; Signal; Transmembrane.
 FT NON_TER 1 1
 FT SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;
 Best Local Similarity 58.9%; Pred. No. 3,2e-67;
 Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

OY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSGNGDMTYARLGKFGTQINSDLTYGGO 60
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 15 AEYNNKNNKLDVYKIKAMHYFSDYDSK----DGDQTYVRFGKGETQINDLDTIGR 69
 61 WEYNNGNNSGADAGTGNKTRLAAGLKYADVGSFDRGNVGVYDALGYTDLMPERG 120
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 70 WESESGKKTSDSSQ---KTRLAAGYKLNKNGSFDRGNLALYDVAMTDMPEERG 126
 121 D-TAVSDPEFVGVRGVATYRNSNFFGLVDGLNFAVOYLKNERDTPARRSGDVGSGIS 179
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 127 DSAQTDFNFMKRRASGLATYRNTDFGLVDGLTLQYOGKNEGREAKQNGDVGTSLS 186
 180 YEYEG--FGIYGAGADRTNLQEAQPLNGCKKAEQMATGLKYDANNITLYLANGETRNA 237
 ||||| :||:|||| :||| : :||| ||||| |||||
 OY 187 YDFGSDFAVSAAYTSSDRTDNLARQGSKAEMATGLKYDANNITLYLANGETRNA 246
 238 TPITNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKAKVEGIGDVLVNYFEV 297
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 247 TPI-----SGGFANKAQNFEAVAOYQDFGLRPSLGYLSGKIDEGVGSDDLNYIDV 300
 298 GATYFNKNNMSTYVDYIINQIDSDNKLGVGSD 330
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 301 GLTYFNKNNMNAFVDYKINQKSDNKLGINDD 333

RESULT 7
 O9RAW2 PRELIMINARY; PRT; 333 AA.
 AC O9RAW2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE PHOSPHATE PORIN (FRAGMENT).
 GN PHO.
 OS Klebsiella pneumoniae subsp. rhinoscleromatis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OX NCBI_TaxId=39831;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NC05046;
 MEDLINE-20023050; PubMed-10555350;
 RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,

RA Kemp D.J.;
 RT "Phylogenetic evidence for reclassification of Calymmatobacterium
 RT granulomatis as Klebsiella granulomatis comb. nov.";
 RT Int. J. Syst. Bacteriol. 49:1695-1700(1999).
 CC -1- SUBUNIT: HOMOTIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 (BY SIMILARITY).
 CC EMBL: AF009230; AAD21518.1; -.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
 KM Outer membrane; Porin; Signal; Transmembrane.
 FT NON_TER 1 1
 FT SEQUENCE 333 AA; 36839 MW; 68AF347207770ABA CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 333;
 Best Local Similarity 58.9%; Pred. No. 3,2e-67;
 Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

OY 1 AEIYNKGNKVDLYGKAVGLHYFSKNGENSGNGDMTYARLGKFGTQINSDLTYGGO 60
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 15 AEYNNKNNKLDVYKIKAMHYFSDYDSK----DGDQTYVRFGKGETQINDLDTIGR 69
 61 WEYNNGNNSGADAGTGNKTRLAAGLKYADVGSFDRGNVGVYDALGYTDLMPERG 120
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 70 WESESGKKTSDSSQ---KTRLAAGYKLNKNGSFDRGNLALYDVAMTDMPEERG 126
 121 D-TAVSDPEFVGVRGVATYRNSNFFGLVDGLNFAVOYLKNERDTPARRSGDVGSGIS 179
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 127 DSAQTDFNFMKRRASGLATYRNTDFGLVDGLTLQYOGKNEGREAKQNGDVGTSLS 186
 180 YEYEG--FGIYGAGADRTNLQEAQPLNGCKKAEQMATGLKYDANNITLYLANGETRNA 237
 ||||| :||:|||| :||| : :||| ||||| |||||
 OY 187 YDFGSDFAVSAAYTSSDRTDNLARQGSKAEMATGLKYDANNITLYLANGETRNA 246
 238 TPITNKFTNTSGFANKTQDVLVAQYQDFGLRPSIATYKSKAKVEGIGDVLVNYFEV 297
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 247 TPI-----SGGFANKAQNFEAVAOYQDFGLRPSLGYLSGKIDEGVGSDDLNYIDV 300
 298 GATYFNKNNMSTYVDYIINQIDSDNKLGVGSD 330
 ||||| :||:|||| :||| : :||| ||||| |||||
 DB 301 GLTYFNKNNMNAFVDYKINQKSDNKLGINDD 333

RESULT 8
 O9K3E7 PRELIMINARY; PRT; 363 AA.
 AC O9K3E7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
 DE OUTER MEMBRANE PROTEIN FX PRECURSOR.
 GN OMPX.
 OS Salmonella enterica subsp. enterica serovar Minnesota.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxId=70803;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF111;
 RA Zimmermann H., Wassenaar T.M., Laubenheimer-Preusse H., Petry F.,
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y15843; CAB96614.1; -.
 DR InterPro: IPR001702; -.
 DR Pfam: PF00267; Gram-ve.porins; 1.
 DR PRINTS: PR00182; ECOLNEIPORIN.
 DR PROSITE: PS00576; GRAM_NEG_PORIN; UNKNOWN_1.
 KM Signal.
 FT SIGNAL 1 22 POTENTIAL.

FT CHAIN 23 363 OUTER MEMBRANE PROTEIN FX.
SQ SEQUENCE 363 AA: 40090 MW: F5058DEDEA516859 CRC64;

Query Match 56.5%; Score 1022.5; DB 2; Length 363;
Best Local Similarity 57.3%; Pred. No. 3.6e-67;
Matches 199; Conservative 54; Mismatches 81; Indels 13; Gaps 8;

QY 1 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 23 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 78
QY 61 WEYNFOGNNSEGADQGTGKTRLAFLAGLKYADVSFDRYRNGVYDALGYTMDLPEFG 120
DB 79 WEYRKADAEQ-EGQNSLVRLAFGLKYAEVSGIDYGRNRTGYDVESYTDMAFYEG 137
QY 121 DT---AYSDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNREDTARRS-NGDVG 176
DB 138 ETWGGAYTDNMTSRAGGLTYRNSDFGLVDGLSFGIYQGNQDNHISNGNDGVCY 197
QY 177 STSYEEGFGIYAGVADRTNLQEAQPLGNGKKAQMATGLKYDANNITYLAANGETR 236
DB 198 TMAVEFDGFGVAAVNSKRTNDQODRD-CNGDRAESMAVGAKEYDANNITYLAAYAE 256
QY 237 ATPITNKFTNTSGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGI-GDVLVNF 295
DB 257 MSIVETVTDYEMANKQTLQEVVAOYQDFGLRPSIATYKSKAKVEGI-GDVLVNF 316
QY 296 EVGATYFNKNNSTYVDYITINOIDSQNLG--VGSDDYAVGVGYO 340
DB 317 QGATYFNKNNSTYVDYITINOIDSQNLG--VGSDDYAVGVGYO 363
RESULT 9
Q9RS59 PRELIMINARY: PRT: 333 AA.
AC Q9RS59: 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 16, Last annotation update)
DE PHOSPHATE PORIN (FRAGMENT).
GN PHOE.
OS Bacteriobacterium granulomatis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Calymmatobacterium.
OX NCBI_TaxID=39824;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20023050; PubMed=10555350;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
Kemp D.J.,
"Phylogenetic evidence for reclassification of Calymmatobacterium
granulomatis as Klebsiella granulomatis comb. nov.";
RL Int. J. Syst. Bacteriol. 49:1695-1700(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
(BY SIMILARITY).
CC EMBL: AF009231; AAD21519.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 333
SQ SEQUENCE 333 AA: 36786 MW: 7CEA2F68B998B0A CRC64;

Query Match 56.5%; Score 1021.5; DB 2; Length 333;
Best Local Similarity 58.9%; Pred. No. 3.8e-67;
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 15 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 69
QY 61 WEYNFOGNNSEGADQGTGKTRLAFLAGLKYADVSFDRYRNGVYDALGYTMDLPEFG 120
DB 70 WESEFGKNTSESSQ---KTRLAFLAGVTKLKNYSFDRYGRNLGALVDVEMTMDPEFG 126
QY 121 D-TAYSDDFEFGVGVATYRNSNFEGLVDGLNFAVOYLGNREDTARRSNGGVCSS 179
DB 127 DSSAQDNMTKTRASGLATYRNTDFGLVDGLTQYQGNKGCBAKQNGDVGTSLS 186
QY 180 YEEG--FGIYAGVADRTNLQEAQPLGNGKKAQMATGLKYDANNITYLAANGETR 237
DB 187 YDFGSGDFAVSAATYSSDRINDONLARQGSKAEMATGLKYDANNITYLATYSETRK 246
QY 238 TPTNKFTNTSGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGIQDVLVNF 297
DB 247 TPT-----SGFANKQDVLVAOYQDFGLRPSIATYKSKAKVEGIQDVLVNF 300
QY 298 GATYFNKNNSTYVDYITINOIDSQNLGVSDD 330
DB 301 GLTYFNKNNSTYVDYITINOIDSQNLGVSDD 333
RESULT 10
Q9S613 PRELIMINARY: PRT: 333 AA.
AC Q9S613: 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 16, Last annotation update)
DE OUTER MEMBRANE PHOSPHATE PORIN PRECURSOR (FRAGMENT).
GN PHOE.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=NCTC9633T;
RA Carter J.S., Bowden F.J., Bastian I., Myers G.M., Sriprakash K.S.,
Kemp D.J.,
"Phylogenetic evidence for reclassification of Calymmatobacterium
granulomatis as Klebsiella granulomatis comb. nov.";
RL Int. J. Syst. Bacteriol. 0:0-0(1999).
CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
(BY SIMILARITY).
CC EMBL: AF064793; AAD11630.1; -.
DR HSSP: P02932; 1PHO.
DR InterPro: IPR001702; -.
DR Pfam: PF00267; Gram-ve_porins; 1.
DR PRINTS: PR00182; ECOLEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.
KW Outer membrane; Porin; Signal; Transmembrane.
FT NON_TER 1
FT NON_TER 333
SQ SEQUENCE 333 AA: 36855 MW: 68AF32C5077705EA CRC64;

Query Match 56.4%; Score 1019.5; DB 2; Length 333;
Best Local Similarity 58.9%; Pred. No. 5.3e-67;
Matches 196; Conservative 41; Mismatches 79; Indels 17; Gaps 5;

QY 1 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 60
DB 15 AEIYNKDKKKVLYGKAVGLHYFSKNGENSGNGDMTYARLGFGEQINSDLTGCG 69
QY 61 WEYNFOGNNSEGADQGTGKTRLAFLAGLKYADVSFDRYRNGVYDALGYTMDLPEFG 120
DB 70 WESEFGKNTSESSQ---KTRLAFLAGVTKLKNYSFDRYGRNLGALVDVEMTMDPEFG 126

QY	121	D-IAYSDDFEYGRVGYATYRNSFFELVUGLFAVOYLCKNERDTPARNSGCGGGGSS	179
Db	127	DSAQOTDNFMFKRMSGATYRNTDFEELVUGLDTLTOCKNKGREGKKGCGGVETSS	186
QY	180	YEEAGS--FGIAGVAGADRTNLOEAPGLGNGSKRAEQMATLKTADANNIYLAANYGERTNA	237
Db	187	YDFGSGDFAVSAATTSDDRTNDQULRLRGGSKRAEMATLKTADANNIYLATYSETRKM	246
QY	238	TPITNKFNTSGFANKQTQDVLVLAQYQDFGLRPSIAYYTKSKAADVEGICDVLVNFEX	297
Db	247	TPI-----SGCFANKQNFQFAVQYQDFGLRPSLGYLTSKGDIEGVSEDDLNVNIYD	300
QY	298	GATYFNKKNMSTYYDIINQIDSNKTKGVGSSD	330
Db	301	GLTYFFNKNMNAFVDYKINQKSNKTKGINDD	333

SEQ	SEQUENCE	360 AA;	39665 MW;	B48A51685965771 CRC64;
RESULT	11			
Q9F889				
ID	Q9F889	PRELIMINARY;	PRT;	360 AA.
AC	Q9F889;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)			
DT	01-MAR-2001 (TREMBLrel. 16, last annotation update)			
DE	OUTER MEMBRANE PROTEIN F. PRECURSOR (FRAGMENT).			
GN	OMP.			
OS	Salmonella typhi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Salmonella.			
OX	NCBI_TaxID=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=9, 12, (VI);D;			
RA	Mata-Gonzalez M.T., Pelayo R., Isidasi A., Montoliu L.,			
RA	Ortiz-Navarrete V.F.;			
RT	"OmpF porin is expressed in Salmonella typhi.";			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBD databases.			
DR	EMBL; AF251685; AAC09474.1; .			
KW	Signal.			
FT	SIGNAL.	1	22	POTENTIAL.
FT	NON_TER	360	360	

Query Match	56.28;	Score 1017.5;	DB 2;	Length 360;
Best Local Similarity	57.68;	Pred. No. 8.3e-67;		
Matches 198; Conservative	54;	Mismatches 79;	Indels 13;	Gaps 8;

[illegible]

RESULT	12			
087753				
ID	087753	PRELIMINARY:	PRT:	359 AA.
AC	087753;			
DT	01-NOV-1998 (Tremblrel, 08, Created)			
DT	01-NOV-1998 (Tremblrel, 08, Last sequence update)			
DT	01-MAR-2001 (Tremblrel, 16, Last annotation update)			
DE	OMPX35 PORIN PRECURSOR.			
GN	OMPX35.			
OS	Klebsiella pneumoniae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Klebsiella.			
OX	NCBI_TaxID=573;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KT755;			
RA	Hernandez-Alles S.;			
RL	Thesis (1998), Universitat de les Illes Balears, Valencia, Spain			
DR	EMBL; AJ011501; CAA0965.1; -.			
DR	HSSP; P02931; IGFN.			
DR	InterPro; IPR001702; -.			
DR	Pfam; PF00267; Gram-Ve_porins; 1.			
DR	PRINTS; PR00182; ECOLNEIPORIN.			
KW	Signal.			
FT	SIGNAL.	1	22	POTENTIAL.
FT	CHAIN	23	352	
SQ	SEQUENCE	359 AA;	39510 MW;	12D51834005B5EBB CRC64;

Query Match	51.7%;	Score 934.5;	DB 2;	Length 359;
Best Local Similarity	55.3%;	Pred. No. 9.6e-61;		
Matches 194;	Conservative 42;	Mismatches 90;	Indels 25;	Gaps 11;

```

Qy      1 AALYKDGKRDVLYKAVGLYFESKQNGENSYGNGDMTARLGFGEQIINDLGYGO 60
      |||||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      23 AELIYKNGKRLDLYFYKMGVHWMT--NGDS---SDTPTARLGLGEQIINQLGYGO 78
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      61 WEYNQGNNSGADQOGCNKTRLAFAGLKYADVSPFYGRNYGVYDALGYDTMLPEFG 120
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
Db      79 WEYNNDASVBS--SQT--TKTRLAFAGLKAGEYSEFDYGRNYGALIDVEAATDMLYEMG 135
      |||||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      121 D--TAYSDPEFYRVGVATYTRNSNFEGLVDLNTAFVLYLAKNERHDTA--RRSNDGVGSI 178
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Db      136 DGMNTDNTMGRCTGVSATYTRNSDPEGLVDLSPALYOGKNDHRAIRQNDGFSTAA 195
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      179 SYEYE--GFGIVAGYAGADRNLQEAOPLGNGKKAQWMTGLKYDANNIYAANYGETRNA 237
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Db      196 TAFADGIALSGYSSNRSYDQKAD--GNGDKAEAMTASAKYDANNIYAAMVSYTMM 253
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      238 TPTLNKFTNSGFANKTDVLLVAYOYDFBLRPSIATYTSKADVGI-----GPDVLY 292
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Db      254 TP----EEDNHFACKTONFEAVVOYQDFGLRPSIGVYQTKGADLQSRAGFSGGADLV 308
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      293 NTFEGATYFENKKNSTYVDYLIINDSDN---KLVGSSDPTVAVLYVOF 340
      ||::||::|||::|||::|||::|||::|||::|||::|||::|||
Db      309 KYLEGTWYFENKKNMNYAAKFNQLDNDNTKKAAGATDQAAVGLYVOF 359
      |||||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 13		
Q9EXH8		
ID	Q9EXH8	PRELIMINARY; PRT; 315 AA.
AC	Q9EXH8;	
DT	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)	
DE	OMPR35 PROTEIN PRECURSOR.	
GN	OMPR35.	
OS	Klebsiella pneumoniae.	
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
OC	Klebsiella.	
OX	NCBI_TaxID=573;	
RP	[1]	
RP	SEQUENCE FROM N.A.	

RC	STRAIN-103624;					
RA	Crowley B., Benedi V.J., Domenech-Sanchez A.;					
RT	Expression of altered OmpK35 porin and SHV-2 beta-lactamase in					
RT	<i>Klebsiella pneumoniae</i> results in cephalosporins and carbapenems					
RT	resistance";					
RL	Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AJ303057; CAC18650.1; --.					
KW	Signal.					
FT	SIGNAL.					
FT	1	22	POTENTIAL.			
FT	CHAIN	23	315	OMPK35 PROTEIN.		
SQ	SEQUENCE	315 AA;	34420 MM;	F45748383475D8CA	CRC64;	

Query Match	45.1%	Score 815;	DB 2;	length 315;
Best Local Similarity	55.6%	Pred. No. 4	3e-52;	
Matches 170;	Conservative 37;	Mismatches 77;	Indels 22;	Gaps 10;

[illegible]

Query Match Similarity	42.6%	Score 770;	DB 2;	Length 315;
Best Local Similarity	55.8%	Pred. NO. 8.4e-49;		
Matches 159; Conservative	33;	Mismatches 67;	Indels 26;	Gaps 8.

QY	1	AEIYNKOCNKVDLKGAKVAHGFYFKSGKNENYSGNGDMWYARLFEKKEQOINSQJLTGQ	60
		: : : : : : : : : : : : : : : : :	
Db	22	AEIYNKOCNKLDLFGRAAKLFLFKQSN-----DATYRIKFEKETHKNSQJLTGQ	74
QY	61	WEYNFGNNSGADAOQGNKTRLAPAGIKYADVGSFDYGRNYGVYDALGYTDMLPBGG	120
		: : : : : : : : : : : : : : : : :	
Db	75	WEYIAANMTE--SOGDDGNKTRLFGAGIKLADYCSFDYGRNYGVYVYLSTYDMLPFBGG	133
QY	121	DT--AYSDFEFGRGAGVATYRNSNPFGLVDGIANVAYOLGKNE-----RDTARRSNGD	172
		: : : : : : : : : : : : : : : : : : : : : :	
Db	134	DSMAATDNYMTRARSTGLATYRNDNPFGLYBELKKAADYQGNNEGCDYTRNRSIQKANGD	199
QY	173	GVGSGISYE-YEGFGI--VGAYGADRTNIQ--EAQPLGNGKAKQMATGLKYDANNIYA	228
		: : : : : : : : : : : : : : : : : : : : :	
Db	194	GFGLSFYNOQNGSGVSAAGAYSSSNRTEGQKELVNSAAGKAQAMATGLKYDANQAYIA	253
QY	229	ANYETRRAPITNKFYTSFGAKTODVLVAYOJDFEGRPSI	273
		: : : : : : : : : : : : : : : : : : : : : :	
Db	254	AMYGTELTMYTKRL-----IAKQTVNELVAYOJDFEENGIRPSL	292

Query Match	23.7%	Score	428.5	DB 2	Length	130			
Best Local	59.9%	Pred. No.	2.2e-24						
Matches	82	Conservative	15	Mismatches	33	Indels	7	Gaps	2

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:50:50 ; Search time 86.32 Seconds
(Without alignments)
204.374 Million cell updates/sec

Title: US-09-490-291-2

Sequence: 1 MRGSHHHHHHSGMASGRGL.....YGGLSGQSTSGIRPAKLN 291

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

A.Geneseq.0601:*

- 1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1036	69.8	606	16	AAAR9053 Spider dragline va
2	1036	69.8	606	20	AAAY40100 Polymer of an anal
3	1031	69.4	646	18	AAAM27178 Nephlia clavipes s
4	1031	68.4	651	20	AAAY40097 Spider silk protei
5	1015	68.4	718	19	AAAS33346 Nephlia clavipes s
6	1015	68.4	718	21	AAAY59070 N. clavipes spider
7	1001	67.4	718	12	AAAR14308 Spider dragline va
8	984	66.3	604	16	AAAR9057 Spider dragline va
9	976	65.7	606	16	AAAR9055 Polymer of an anal
10	976	65.7	606	20	AAAY40101 Polymer of an anal
11	976	65.7	606	20	AAAY40102 Polymer of an anal

12	677.5	45.6	615	20	AAAY40099 Spider silk protei
13	674	45.4	831	16	AAAR80168 PMISSI Misp spider
14	610.5	41.1	641	20	AAAT28863 Epstein Barr Virus
15	610.5	41.1	641	21	AAAY5856 Epstein Barr Virus
16	593.5	40.0	738	19	AAAS56165 New DNA sequence 1
17	591.5	39.8	261	19	AAAW79137 FLGA Gly-ala inser
18	561	37.8	235	17	AAAO5704 Glycine-rich repea
19	561	37.8	235	19	AAAT79126 Epstein Barr Virus
20	554	37.3	331	20	AAAY40098 Spider silk protei
21	554	37.3	331	12	AAAR14309 Nephlia clavipes s
22	554	37.3	331	19	AAAS33347 N. clavipes spider
23	554	37.3	331	16	AAAY59071 Spider dragline va
24	524	35.3	714	16	AAAR9059 Polymer of an anal
25	524	35.3	714	20	AAAY40103 Polymer of an anal
26	520	35.0	1177	9	AAAB0940 SLP111 protein com
27	520	35.0	1177	11	AAAR05307 SLP IIT (silk-fibr
28	520	35.0	1177	17	AAAR5105 Silk-like protein
29	520	35.0	1177	18	AAAM26342 SLP111-like protein
30	520	35.0	1177	21	AAAY78277 SLP111 amino acid
31	520	35.0	1177	22	AAAB72275 Repetitive protein
32	520	35.0	1177	22	AAAB6395 SLP111 protein seq
33	520	35.0	1178	14	AAAR41007 Silk-like protein
34	520	35.0	1178	19	AAAS3518 Amino acid sequenc
35	507	34.1	1023	19	AAAS3524 Amino acid sequenc
36	507	34.1	1059	9	AAAB29662 SLP14 protein comp
37	507	34.1	1059	14	AAAR41013 SLP4 multimeric pr
38	507	34.1	1059	18	AAAM26348 SLP4 synthetic pro
39	507	34.1	1101	21	AAAY78283 Adhesion protein.
40	503.5	33.9	980	16	AAAR81318 Silk like protein
41	488	32.9	1332	17	AAAR95109 Repetitive protein
42	488	32.9	1332	22	AAAB72279 SLP-C protein sequ
43	488	32.9	1332	22	AAAB3959 SLP-like protein
44	478	32.2	1018	16	AAAW01496 Fibronectin cell b
45	476.5	32.1	1038	17	AAAR95107

ALIGNMENTS

RESULT 1	
ID	AAAR9053 standard; Protein: 606 AA.
AAAR9053:	
AC	17-JAN-1997 (first entry)
XX	
XX	
DE	Spider dragline variant, DP-1A.9 polymer.
XX	
XX	Spider; dragline protein; variant; monomer; polymer;
KW	fibrous forming region; Spidroin 1; Nephlia clavipes; DPL; mamic;
KW	DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
KW	rope; surgical suture; implant; reinforcement; film; coating.
XX	
OS	Synthetic.
XX	
PN	WO9429450-A2.
XX	
PD	22-DEC-1994.
XX	
PF	15-JUN-1994; 94WO-US06689.
XX	
PR	15-JUN-1993; 93US-0077600.
XX	
PA	(DUPO) DU PONT DE NEMOURS & CO E. I.
XX	
PI	Fahnestock SR;
XX	
XX	WPI; 1995-036479/05.
DR	New synthetic variants of spider dragline protein - for making
XX	fibres useful as clothing, surgical silk, plastic reinforcement
PT	etc., also related DNA, vectors and transformed cells
PT	

CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.

XX Sequence 651 AA;

Query Match 68.4% Score 1015; DB 20; Length 651;
Best Local Similarity 59.2% Pred. No. 1.5e-71;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```
QY 11 GSMASGRGLGCGGAGAAAA-----AAAAAGCAGCGGCG 45
  || :|||||
Db 93 gsgagagrg1g9gagaaaaaaggag9gy9g1gngagrg9ggaagaaaggag9gy9 152
  || :|||||
QY 46 GLSGGTGGRGLGCGGAGAAAA-----AGCAGCGGCGGCGTSGRGL 96
  || :|||||
Db 153 g1gsgg-agrg1g9gagaaaaaaggag9gy9g1gngag9gy9g1gsgg-agrg1 210
  || :|||||
QY 97 GGCGAGAAAA-----AAAGCAGCGGCGGCGTSGRGLG 135
  || :|||||
Db 211 g9gagaaaaaaggag9gy9g1gngag9gagaaaaaggag9gy9g1gsgg----agrg 266
  || :|||||
QY 136 GCGAGAAAA-----AGCAGCGGCGGCGTSGRGLGCGGAGAAAA-- 182
  || :|||||
Db 267 gsgagaaaaaggag9gy9g1gngag9gy9g1gsgg-agrg1g9gagaaaaaggag 325
  || :|||||
QY 183 -----AAAAAGCAGCGGCGGCGTSGRGLGCGGAGAAAA----- 220
  || :|||||
Db 336 g9g1g9gag9gagaaaaaggag9gy9g1gsgg-agrg1g9gagaaaaaggag 384
  || :|||||
QY 221 -----AAAAAGCAGCGGCGGCGTSGRGLGCGGAGAAAA-- 259
  || :|||||
Db 365 g9gy9g1gsgag9gy9gagaaaaaggag9gy9g1gngg-agrg1g9gagaaaaaa 443
  || :|||||
QY 260 -----AAAGCAGCGGCGGCGTSGRGLGCGGAG 279
  || :|||||
Db 444 g9ag9gy9g1gngagrg9gagaaaaaggag9gy9g1gsgg 485
```

RESULT 5

AAW53346
ID AAW53346 standard; Protein: 718 AA.

XX AC AAW53346;

XX DT 06-JUL-1998 (first entry)

XX DE Nephila clavipes spider silk protein.

XX KW Spider: Nephila clavipes; silk protein; tandem repeat; fibre; dragline;

XX KM cocoon; tensile strength; elasticity.

XX OS Nephila clavipes.

XX PN US5728810-A.

XX PD 17-MAR-1998.

XX PF 19-APR-1995; 95US-0425069.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX PR 04-OCT-1994; 94US-0317844.

XX PR 19-APR-1995; 95US-0425069.

XX PA (UYWY-) UNIV WYOMING.

XX PI Himan MB, Lewis RV, Xu M;

XX WPI: 1998-270437/24.

XX DR N-PSDB; AAV23249.

XX PT Recombinant spider silk proteins - useful for making fibres
XX PS Claim 1; Column 29-34; 68pp; English.

CC The present sequence represents a spider silk protein from the present
CC invention. Spider silk proteins, and peptide fragments within the
CC proteins, can be produced and purified independently and can then be
CC mixed and made into fibres that have higher tensile strengths and
CC elasticity than naturally occurring fibres. The fibres can be used in
CC mixed composites. The invention allows the two naturally occurring
CC Nephila clavipes silk proteins to form silk fibres of high tensile strength
CC and elasticity.

XX Sequence 718 AA;

Query Match 68.4% Score 1015; DB 19; Length 718;
Best Local Similarity 59.2% Pred. No. 1.6e-71;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```
QY 11 GSMASGRGLGCGGAGAAAA-----AAAAAGCAGCGGCG 45
  || :|||||
Db 93 gsgagagrg1g9gagaaaaaaggag9gy9g1gngagrg9ggaagaaaggag9gy9 152
  || :|||||
QY 46 GLSGGTGGRGLGCGGAGAAAA-----AGCAGCGGCGGCGTSGRGL 96
  || :|||||
Db 153 g1gsgg-agrg1g9gagaaaaaaggag9gy9g1gngag9gy9g1gsgg-agrg1 210
  || :|||||
QY 97 GGCGAGAAAA-----AAAGCAGCGGCGGCGTSGRGLG 135
  || :|||||
Db 211 g9gagaaaaaaggag9gy9g1gngag9gagaaaaaggag9gy9g1gsgg----agrg 266
  || :|||||
QY 136 GCGAGAAAA-----AGCAGCGGCGGCGTSGRGLGCGGAGAAAA-- 182
  || :|||||
Db 267 gsgagaaaaaggag9gy9g1gngag9gy9g1gsgg-agrg1g9gagaaaaaggag 325
  || :|||||
QY 183 -----AAAAAGCAGCGGCGGCGTSGRGLGCGGAGAAAA----- 220
  || :|||||
Db 326 g9g1g9gag9gagaaaaaggag9gy9g1gsgg-agrg1g9gagaaaaaggag 384
  || :|||||
QY 221 -----AAAAAGCAGCGGCGGCGTSGRGLGCGGAGAAAA-- 259
  || :|||||
Db 385 g9gy9g1gsgag9gy9gagaaaaaggag9gy9g1gngg-agrg1g9gagaaaaaa 443
  || :|||||
QY 260 -----AAAGCAGCGGCGGCGTSGRGLGCGGAG 279
  || :|||||
Db 444 g9ag9gy9g1gngagrg9gagaaaaaggag9gy9g1gsgg 485
```

RESULT 6

AAV59070
ID AAV59070 standard; Protein: 718 AA.

XX AC AAV59070;

XX DT 08-MAR-2000 (first entry)

XX DE N. clavipes spider silk protein 1.

XX KW Spider silk protein; dragline silk protein; major ampullate gland; fiber.

XX OS Nephila clavipes.

XX PN US5989894-A.

XX PD 23-NOV-1999.

XX PF 04-OCT-1994; 94US-0317844.

XX PR 15-APR-1991; 91US-0684819.

XX PR 20-APR-1990; 90US-0511792.

XX (UYWY-) UNIV WYOMING.
 PA Himman MB, Xu M, Lewis RV;
 PI WPI: 2000-061225/05.
 DR N-PSDB; AA238195.
 XX
 PT Isolated DNA, vector and transformed cell encoding for and useful in
 PT the production of spider silk protein -
 XX
 PS Claim 1; Fig 6A-D; 65pp; English.
 XX
 CC The invention provides isolated cDNA molecules coding for spider silk
 CC proteins. The spider silk proteins are characterized by repeating alpha
 CC and beta regions and optional variable regions. The DNA sequences are
 CC useful in the production of spider silk protein by recombinant DNA
 CC techniques. The recombinant spider silk protein may be used for the
 CC production of fibers. The present sequence represents the spider silk
 CC protein 1, derived from the major ampullate gland of Nephila clavipes.
 CC
 XX Sequence 718 AA;
 SQ

Query Match 68.4%; Score 1015; DB 21; Length 718;
 Best Local Similarity 59.2%; Pred. No. 1.6e-71;
 Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

QY 11 GSMASGRGCLGCGAGAAAA-----AAAAAGAGGCGGCG 45
 DB 93 gsggagagrgl1gsggagagagagagagrgl1gngagrggagagagagagagrgg 152
 QY 46 GLGSGCTSGRGCLGCGAGAAAA-----AGGAGCGGCGGCGGCGTSGRGCL 96
 DB 153 glsgag-agrggl1gsggagagagagagagrgg1gsggagagrgg1gsgg-agrggl 210
 QY 97 GCGGAGAAAA-----AAAAAGAGGCGGCGGCGTSGRGCLG 135
 DB 211 gsggagagagagagagrgg1gsggagagagagagagagrgg1gsgg-----agrg 266
 QY 136 GCGGAGAAAA-----AGGAGCGGCGGCGTSGRGCLGCGAGAAAA----- 182
 DB 267 gsggagagagagagagrgg1gsggagagagagagrgg1gsgg-agrggl1gsggagagagag 325
 QY 183 -----AAAAAGAGGCGGCGGCGTSGRGCLGCGAGAAAA----- 220
 DB 326 gsggagagagagagagagrgg1gsggagagagagagagrgg1gsgg-----agrg 384
 QY 221 -----AAAAAGAGGCGGCGGCGTSGRGCLGCGAGAAAA----- 259
 DB 385 gsggag1gsggagrgg1gsggagagagagagrgg1gngg-agrggl1gsggagagagagag 443
 QY 260 -----AAAAAGAGGCGGCGGCGTSGRG 279
 DB 444 gsggagrgg1gnggagrggagagagagagrgg1gsgg 485

RESULT 7
 AAR14308
 ID AAR14308 standard; Protein: 718 AA.
 AC AAR14308;
 XX
 XX 15-JAN-1992 (first entry)
 DE N.clavipes dragline silk protein-1.
 XX
 KM protein superfibre: major ampullate silk; orb web spider.
 XX Nephila clavipes.
 OS
 XX EP452925-A.
 PN
 XX

PD 23-OCT-1991.
 XX
 XX 18-APR-1991; 91EP-0106217.
 PE
 XX 20-APR-1990; 90US-0511792.
 PR
 XX (UYWY-) UNIV OF WYOMING.
 PA
 XX Lewis RV, Xu M, Himman M;
 PI WPI: 1991-312199/43.
 DR N-PSDB; AAQ14183.
 DR
 XX
 PT DNA encoding spider silk protein-1 and 2 and variants - isolated
 PT from Nephila clavipes, for prodn. of spider silk protein and
 PT fibres having desired characteristics
 XX
 PS Claim 15; Page 23; 48pp; English.
 XX
 CC The spider silk protein contains a basic 34 amino acid repeat. The
 CC repeat itself contains 3 regions. The first comprises 0-9 amino
 CC acids with a sequence AGR(GGX)2. This region is not highly
 CC conserved. The second region has a sequence GAG(A)x which is highly
 CC conserved and is 8-10 amino acids long. The third segment is (GGX)5
 CC and is 15 amino acids long and is very highly conserved. In most
 CC cases X is A,Q, Y or L. Removal of the poly-(Ala) segments results
 CC in a silk having lower elasticity.
 CC
 XX Sequence 718 AA;
 SQ

Query Match 67.4%; Score 1001; DB 12; Length 718;
 Best Local Similarity 58.7%; Pred. No. 1.9e-70;
 Matches 236; Conservative 8; Mismatches 16; Indels 142; Gaps 13;

QY 11 GSMASGRGCLGCGAGAAAA-----AAAAAGAGGCGGCG 45
 DB 93 gsggagrgg1gsggagagagagagagrgg1gngagrggagagagagagagrgg 152
 QY 46 GLGSGCTSGRGCLGCGAGAAAA-----AGGAGCGGCGGCGGCGTSGRGCL 96
 DB 153 glsgag-agrggl1gsggagagagagagagrgg1gsggagagrgg1gsgg-agrggl 210
 QY 97 GCGGAGAAAA-----AAAAAGAGGCGGCGGCGTSGRGCLG 135
 DB 211 gsggagagagagagagrgg1gsggagagagagagagagrgg1gsgg-----agrg 266
 QY 136 GCGGAGAAAA-----AGGAGCGGCGGCGTSGRGCLGCGAGAAAA----- 182
 DB 267 gsggagagagagagagrgg1gsggagagagagagrgg1gsgg-agrggl1gsggagagagag 325
 QY 183 -----AAAAAGAGGCGGCGGCGTSGRGCLGCGAGAAAA----- 220
 DB 326 gsgg1gsggagagagagagagagrgg1gsggagagagagagagrgg1gsgg-----agrg 384
 QY 221 -----AAAAAGAGGCGGCGGCGTSGRGCLGCGAGAAAA----- 259
 DB 385 gsggag1gsggagrgg1gsggagagagagagrgg1gngg-agrggl1gsggagagagagag 443
 QY 260 -----AAAAAGAGGCGGCGGCGTSGRG 279
 DB 444 gsggagrgg1gnggagrggagagagagagrgg1gsgg 485

RESULT 8
 AAR9057
 ID AAR9057 standard; Protein: 604 AA.
 AC AAR9057;
 XX
 XX 17-JAN-1997 (first entry)
 DE Spider dragline variant, DP-1B.16 polymer.
 XX

XX Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephrila clavipes; Dp1; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX Synthetic.
 OS
 PN M09429450-A2.
 PD 22-DEC-1994.
 XX 15-JUN-1994; 94WO-US06689.
 PF 15-JUN-1993; 93US-0077600.
 PR 15-JUN-1993; 93US-0077600.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX Fahnestock SR;
 PI
 XX WPI; 1995-036479/05.
 DR
 XX New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS
 XX Disclosure; Page 106-108; 168pp; English.
 XX This sequence represents a synthetic spider dragline variant polymer,
 CC DP-1B.16. The sequence of the DP-1B.16 monomer is given in AAR99056.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephrila clavipes. The
 CC DNA sequence encoding the monomer may be used in the recombinant
 CC production of the variant protein in a recombinant host, e.g. E. coli
 CC or Bacillus subtilis. Synthetic analogues of Dp1 were designed to mimic
 CC the repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. This monomer exhibits all of the
 CC regularities of (1)-(5) below. In addition, it exhibits a regularity of
 CC the natural sequence which is not shared by DP-1A, namely that a repeat
 CC in which both GYG and GRG are deleted is generally preceded by a repeat
 CC lacking the entire poly-alanine repeat, with one intervening repeat.
 CC The sequence of DP-1B matches the natural sequence more closely over
 CC a more extended segment than does DP-1A. The individual repeats differ
 CC from the consensus sequence given in AAW06201 according to the pattern:
 CC (1) the poly-alanine sequence in length from 0-7
 CC residues; (2) when the entire poly-alanine sequence is deleted,
 CC so also is the surrounding sequence encompassing AGRGIGGCGAGAGC;
 CC (3) aside from the poly-alanine sequence, deletions usually
 CC encompass integral multiples of three consecutive residues;
 CC (4) deletion of GYG is generally accompanied by deletion of GRG
 CC in the same sequence; and
 CC (5) a repeat in which the entire poly-alanine sequence is
 CC deleted is generally preceded by a repeat containing six alanine
 CC residues.
 CC The proteins may be used to produce fibres of high tensile strength and
 CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
 CC for implants, plastic reinforcements, films, coatings, etc.
 XX
 XX Sequence 604 AA;
 XX
 Query Match 66.3%; Score 984; DB 16; Length 604;
 Best Local Similarity 60.2%; Pred. No. 3.4e-69;
 Matches 231; Conservative 4; Mismatches 13; Indels 136; Gaps 13;
 OY 11 GSMASGRGLGCGGCAAAA-----AAAAAGAGGCGTGGGCGS 49
 II :|||||
 DB 13 gsggagrg1gsggagaaagagaggg1gsggagggagagaaagagggg1gsg 72
 OY 50 OCTSRGGLGSGGAGAAAAAGAGGCGGTGSGGT-----SGRGGCG 97
 II :|||||
 DB 73 gg-----agrgggag-----aaaaagagggg1gsggagggg1gsggagrg1g 124

OY 98 GCGAGAAA-----AAAAAGAGGCGTGGGCGTSGRGGLG 136
 II :|||||
 DB 125 gsggagaaagagaggg1gsggagggagagaaagagggg1gsgg-----agrgg 180
 OY 137 OGAGAAAAAGAGGCGGTGSGGT-----SGRGGLGCGGAGAAAA-- 182
 III :|||||
 DB 181 gga-----aaaaagagggg1gsggagggg1gsggagrg1gsggagagaaaaa 236
 OY 183 -----AAAAAGAGGCGTGGGCGTSGRGGLGCGGCAAAAAA 223
 II :|||||
 DB 237 aggaggg1gsggagggagagaaagaggg1gsgg-----agrgggagagaaaaag 292
 OY 224 AAAG-----GAGGCGTGGGCGTSGRGGLGCGGAGAAAA-- 258
 II :|||||
 DB 293 gaggggg1gsggagggg1gsggagrg1gsggagagagaaagaggg1gsggag 351
 OY 259 ---AAAAAGAGGCGGTGSGGT-----SGRGGLGCGGAGAAAA-- 258
 II :|||||
 DB 352 gagaagagagggg1gsgg 375
 RESULT 9
 AAR99055
 ID AAR99055 standard; Protein: 606 AA.
 XX
 AC AAR99055;
 XX
 DT 17-JAN-1997 (first entry)
 DT
 XX Spider dragline variant, DP-1B.9 polymer.
 DE
 XX Spider: dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephrila clavipes; Dp1; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX Synthetic.
 OS
 PN M09429450-A2.
 PD 22-DEC-1994.
 XX 15-JUN-1994; 94WO-US06689.
 PF 15-JUN-1993; 93US-0077600.
 PR 15-JUN-1993; 93US-0077600.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX Fahnestock SR;
 PI
 XX WPI; 1995-036479/05.
 DR
 XX New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS
 XX Disclosure; Page 88-90; 168pp; English.
 XX This sequence represents a synthetic spider dragline variant polymer,
 CC DP-1B.9. The sequence of the DP-1B.9 monomer is given in AAR99054.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephrila clavipes. The
 CC DNA sequence encoding the monomer may be used in the recombinant
 CC production of the variant protein in a recombinant host, e.g. E. coli
 CC or Bacillus subtilis. Synthetic analogues of Dp1 were designed to mimic
 CC the repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. This monomer exhibits all of the
 CC regularities of (1)-(5) below. In addition, it exhibits a regularity of
 CC the natural sequence which is not shared by DP-1A, namely that a repeat
 CC in which both GYG and GRG are deleted is generally preceded by a repeat
 CC lacking the entire poly-alanine repeat, with one intervening repeat.

[illegible]

Dd	287	aaagagaggggylgsqaggggyglgsqagrgyglsggaagaataaaggaggglygs	346
OY	259	-----AAAAAGAGGCGGGGLGSQG	279
Dd	347	qgaggaagaagaagaggggylgsqg	375
 RESULT 12			
xx	AAy40099	standard; protein:	615 AA.
xx	AAy40099:		
xx	19-NOV-1999	(first entry)	
xx		Spider silk protein spiderine minor 1.	
xx		Spider silk protein: spiderine minor 1; cosmetic; make-up;	
KW		dermatological compositions; hair care; skin care; sunscreen;	
KM		homone; moisturizer; skin disorder; skin disorder.	
xx		Nephila clavipes.	
OS		FR2774588-AI.	
PN		13-AUG-1999.	
xx		11-FEB-1998;	98FR-0001614.
PF		11-FEB-1998;	98FR-0001614.
xx		11-FEB-1998;	98FR-0001614.
PR		(OREA) L'OREAL SA.	
xx		Philippe M, Garson JC, Arraudau JP;	
PI		WPI, 1999-510729/43.	
DR		Cosmetic or dermatological composition containing spider silk protein,	
PT		for hair or skin care, in make-up or sunscreens	-
xx		Claim 7; Fig 3; 32p; French.	
PS		The present sequence represents the natural spider silk protein	
xx		spidroine minor 1. The protein improves the moisturizing/softening	
CC		action of the compositions. The protein, and its fragments are used	
CC		in cosmetic or dermatological compositions. These compositions have	
CC		use as hair or skin care products, and make-up or sunscreens.	
CC		As the protein is a good, persistent film-formers on the skin	
CC		of low surface density, it can be used for delivery of active	
CC		agents that are generally difficult to administer, e.g. vitamins,	
CC		hormones, moisturizers or agents for treating disorders of the	
CC		skin and hair.	
xx		Sequence 615 AA:	
xx			
SQ			
 Query Match 45.6%; Score 677.5; DB 20; Length 615; Best Local Similarity 47.7%; Pred. No. 1.7e-45; Matches 167; Conservative 10; Mismatches 90; Indels 83; Gaps 13.			
OY	11	GSMASGRGLGOGGA-AAAAAAAAAGAGGCGYG----	GLSGQTSGRGLGQ--- 61
Dd	60	gaggaagggggagagaaaagagaggggylrgagagaaagagaggggggylgs	119
OY	62	--GACAAAAGAGGAGCGYG--GLSGQTSGR-----	GLLGOGAGCAAAAAA 109
Dd	120	ggagagaaaaagagsggggylrgagagaaagagagsggggggylgsagagaa	179
OY	110	AAGCAGCGYG-----GLSGQT-----	SGRGLGOGCAGCAAAAAA 147
Dd	180	agagaggggylrgagagagagaaaragagagaaaaagagaggggylrgagagaa	239

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:47:50 ; Search time 48.42 Seconds

(Without alignments)
121.068 Million cell updates/sec

Title: US-09-490-291-2

1485

Perfect score: 1 MRGSHHHHSGMASGRGL.....YGLGSGTSGIRPAKLN 291

Sequence: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5C_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	68.4	718	1	US-08-425-069-2
2	1015	68.4	718	2	US-08-317-844B-2
3	1015	68.4	747	4	US-09-034-177-3
4	674	45.4	832	1	US-08-209-747-2
5	674	45.4	832	1	US-08-458-298-2
6	593.5	40.0	738	3	US-08-864-038A-3
7	561	37.8	235	2	US-08-529-190B-1
8	554	37.3	595	2	US-08-425-069-4
9	554	37.3	595	2	US-08-317-844B-4
10	520	35.0	1177	1	US-07-609-716-31
11	520	35.0	1177	1	US-08-175-155-29
12	520	35.0	1177	1	US-08-477-509B-64
13	520	35.0	1177	2	US-08-707-237A-35
14	520	35.0	1177	3	US-08-482-085B-64
15	520	35.0	1177	4	US-08-475-411A-31
16	520	35.0	1177	4	US-08-478-029A-31
17	507	34.1	1059	2	US-08-175-155-48
18	507	34.1	1059	2	US-08-707-237A-54
19	507	34.1	1101	3	US-08-477-509B-83
20	507	34.1	1101	3	US-08-482-085B-83
21	488	32.9	1332	1	US-07-609-716-41
22	488	32.9	1332	4	US-08-475-411A-41
23	488	32.9	1332	4	US-08-478-029A-41
24	480.5	32.4	318	4	US-09-060-756-727
25	476.5	32.1	1038	1	US-07-609-716-36
26	476.5	32.1	1038	4	US-08-475-411A-36
27	476.5	32.1	1038	4	US-08-478-029A-36

28	475	32.0	766	1	US-08-175-155-53	Sequence 53, Appl
29	475	32.0	766	1	US-08-477-509B-88	Sequence 88, Appl
30	475	32.0	766	2	US-08-707-237A-61	Sequence 61, Appl
31	475	32.0	766	3	US-08-482-085B-88	Sequence 88, Appl
32	475	32.0	979	1	US-08-477-509B-89	Sequence 89, Appl
33	475	32.0	979	3	US-08-482-085B-89	Sequence 89, Appl
34	475	32.0	1050	1	US-08-175-155-54	Sequence 54, Appl
35	472.5	31.8	784	1	US-07-609-716-48	Sequence 48, Appl
36	472.5	31.8	784	4	US-08-475-411A-48	Sequence 48, Appl
37	472.5	31.8	784	4	US-08-478-029A-48	Sequence 48, Appl
38	470.5	31.7	745	4	US-09-060-756-728	Sequence 728, App
39	469	31.6	870	2	US-09-010-928B-2	Sequence 28, Appl
40	469	31.6	870	2	US-09-010-928B-2	Sequence 2, Appl
41	468.5	31.5	649	1	US-07-609-716-49	Sequence 49, Appl
42	468.5	31.5	649	4	US-08-475-411A-49	Sequence 49, Appl
43	468.5	31.5	649	4	US-08-478-029A-49	Sequence 49, Appl
44	468	31.5	945	1	US-08-089-862-6	Sequence 6, Appl
45	468	31.5	945	1	US-08-587-333-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Hsiao, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8050
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2
Query Match 68.4%; Score 1015; DB 1; Length 718;
Best Local Similarity 59.2%; Pred. No. 2.8e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;
11 GSMASGNGIGGCGAGAAA-----AAAAAGGAGGGGCG 45

```
Db 93 GSOGAGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCGGGG 152
OY 46 GIGSGGTSGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCG 96
Db 153 GIGSGG -AGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCG 210
OY 97 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 135
Db 211 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 266
OY 136 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 182
Db 267 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 325
OY 183 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 220
Db 336 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 384
OY 221 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 259
Db 385 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 443
OY 260 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 279
Db 444 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 485
```

```
RESULT 2
US-08-317-844B-2
; Sequence 2, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317, 844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-317-844B-2
```

```
Query Match 68.4%; Score 1015; DB 2: Length 718;
Best Local Similarity 59.2%; Pred. No. 2.8e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;
OY 11 GSMASGRGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCGGGG 45
Db 93 GSOGAGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCGGGG 152
OY 46 GIGSGGTSGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCG 96
Db 153 GIGSGG -AGRGGLGGGAGAAAAAGAGGCGGCGGAGAAAAAGAGCG 210
OY 97 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 135
Db 211 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 266
OY 136 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 182
Db 267 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 325
OY 183 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 220
Db 336 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 384
OY 221 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 259
Db 385 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 443
OY 260 -----AGAGAGGCGGCGGAGAAAAAGAGGCGGCGGAGAAAAAG 279
Db 444 GGAGAGAAAAAGAGGCGGAGAAAAAGAGGCGGCGGAGAAAAAGAG 485
```

```
RESULT 3
US-09-034-177-3
; Sequence 3, Application US/09034177
; Patent No. 6127146
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,177
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0486 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: GI 1174414
US-09-034-177-3

Query Match 68.4%; Score 1015; DB 4; Length 747;
Best Local Similarity 59.2%; Pred. No. 2.9e-75;
Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

QY 11 GSMASGSGGLGGGAGAAAA-----AAAAAGAGGCGGCG 45
DB 93 GSGGAGGGLGGGAGAAAAAGAGCGGCGGGLGNAGCGCGGAGAAAAAGAGGCGGCG 152
QY 46 GLGSGGTSGRGLGGGAGAAAAAGAGCGGCGGGLGNAGCGCGGAGAAAAAGAGGCGGCG 96
DB 153 GLGSGG-AGRGLGGGAGAAAAAGAGCGGCGGGLGNAGCGCGGAGAAAAAGAGGCGGCG 210
QY 97 GCGGAGAAAA-----AAAGAGCGGCGGGLGNAGCGCGGAGAAAAAGAGGCGGCG 135
DB 211 GCGGAGAAAAAGAGCGGGLGGGAGAGASAAAAAGAGGCGGGLGNAGCGGCGGAG 266
QY 136 GCGGAGAAAA-----AGAGCGGCGGGLGNAGCGGGLGNAGCGGCGGAGAG 182
DB 267 GEGAGAAAAAGAGCGGGLGGGAGAGCGGCGGGLGNAGCGGCGGAGAGAGAGAG 325
QY 183 -----AAAAAGAGCGGCGGGLGNAGCGGGLGNAGCGGAGAAAA----- 220
DB 326 GCGGAGCGAGAGAAAAAGAGCGGCGGGLGNAGCGGGLGNAGCGGAGAAAAAGAGAG 384
QY 221 -----AAAAAGAGCGGCGGGLGNAGCGGGLGNAGCGGCGGAGAAAA-- 259
DB 385 GCGGGLGGGAGGCGGAGAAAAAGAGCGGCGGGLGNAGCGGGLGNAGCGGCGGAGAAAA 443
QY 260 -----AAAAAGAGCGGCGGGLGNAGCGGGLGNAGCGGCGGAGAAAA 279
DB 444 GAGCGGCGGGLGNAGCGGCGGAGAAAAAGAGCGGCGGGLGNAGCGGCGGAG 485

RESULT 4

US-08-209-747-2
Sequence 2, Application US/08209747
Patent No. 5733771

GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.

TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia

COUNTRY: USA
ZIP: 22040-3487

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn, Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/209,747
FILING DATE: 14-MAR-1994

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
HYPOTHEetical: NO

FRAGMENT TYPE: Internal
ORIGINAL SOURCE:

ORGANISM: N. clavipes
TISSUE TYPE: minor ampullate gland

FEATURE:
NAME/KEY: CDS
LOCATION: 1..309

US-08-209-747-2

Query Match 45.4%; Score 674; DB 1; Length 832;
Best Local Similarity 46.4%; Pred. No. 1.1e-47;
Matches 168; Conservative 7; Mismatches 97; Indels 90; Gaps 13;

QY 15 SGRGLGGGAGAAAAAGAGAGCGGCG-----LGSGGTSGRGLG--- 59
DB 314 AGAGAGGCGGAG 373
QY 60 GCGGAGAAAAAG 108
DB 374 GAGAGAAAAAG 433
QY 109 AAAAGAGAGGCGGGLGNAGCGGGLGNAGCGGGLGNAGCGGGLGNAGCGGAGAGAG 156
DB 434 AAA--GAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 491
QY 157 -GCGGLGSGT-----SGRGLG-GCGAGAAAAA-----AAAGAGAGCGGCG 197
DB 492 AGYGCGGCGGAG 551
QY 198 GLGSGT-----SGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
DB 552 GCGGAG 611
QY 236 ---GLGSGT-----GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 286
DB 612 AGAGAGAAAAAG 671
QY 287 AA 288
DB 672 AA 673

RESULT 5

US-08-458-298-2
Sequence 2, Application US/08458298
Patent No. 5756677

GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.

TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia

COUNTRY: USA
ZIP: 22040-3487

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

```

      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/458,298
      FILING DATE: 02-JUN-1995
      CLASSIFICATION: 530
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/209,747
      FILING DATE: 14-MAR-1994
      ATTORNEY/AGENT INFORMATION:
      NAME: Murphy Jr., Gerald M.
      REGISTRATION NUMBER: 28,977
      REFERENCE/DOCKET NUMBER: 1447-104P
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 703-205-8000
      TELEFAX: 703-205-8050
      INFORMATION FOR SEQ ID NO: 2:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 832 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      HYPOTHEetical: NO
      FRAGMENT TYPE: internal
      ORIGINAL SOURCE:
      ORGANISM: N. clavipes
      TISSUE TYPE: minor ampullate gland
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..309
      IS-08-458-298-2

```

Query Match	45.48;	Score 674;	DB 1;	Length 832;
Best Local Similarity	46.48;	Pred. NO. 1.1e-47;		
Matches 168; Conservative	7;	Mismatches 97;	Indels 90;	Gaps 13;

QY	15	SGRGGTGGGAGAAAAAAAGAGCGGTGG-----TGSGGTSGRGLG---	59
Db	314	AGAGGAGGTGRRGAGAGGAAAGAGAGGTGCGGCGTGAGAGAAAAAGAGAGGTGRCG	3737
QY	60	GQGAGAAAAAAAGGAGGCGGTGCTGSGGT-----SGRGLGGGAGAAAAAA	108
Db	374	GAGAGAAAGAGAGAAAGAGAGGTGGGCGTGAGAGAGAAAAAGAGAGGAGGTGRRAGAG	4333
QY	109	AAAAAGAGCGGTGGLGSGG-----TSGRGLGGGAGAGAAAAAAAGAGAGCG-	156
Db	434	AAA--GAGAGGTGGGCGGTGAGAGGAAAAAATGAGGAGGTGRRAGAGAGAAAGAGTGG	4919
QY	157	-GYGGLGSGGT-----SGRGLG--GGGAGAAAAA-----AAAAGGCGGTG	197
Db	492	AGTGGGCGGTGAGAGAGAAAAAGAGAGAGTGTGRRAGAGAGAAAGAGAGAGAGAGGTG	5511
QY	198	GLGSGGT-----SGRGLGGGAGAAAAAAAGAGAGCGGTG	235
Db	552	GQGGYGAGARAGAAAAAGAGGAGAGTGTGRRAGAGAGAAAAAGAGAGGTGCGGCGTG	6111
QY	236	---GLG-----GCTSGRGLG--GQGAGAAAAAAAGAGAGCGGTGCTGSGGTGTTTTP	286
Db	612	AGAGAGAAAAAGAGSGGAGGTGRRAGAGAAAGAGAAAGAGAGAGGTGGGCGTGAGAGAAA	6711
QY	287	AA 288	
Db	672	AA 673	

RESULT, 6
US-08-864-038A-3
; Sequence 3, Application US/08864038A
; Patent No. 6001592
; GENERAL INFORMATION:
; APPLICANT: Kunio NAKASHIMA et al

TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
 TITLE OF INVENTION: CONTRAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
 TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
 TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
 TO SAID POLYPEPTIDE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: 812-5 Hlrano
 STREET: Ishinden
 City: Tsu-city
 STATE: Mie-prefecture
 COUNTRY: JAPAN
 Zip: 514-01
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 MB storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Microsoft windows 95
 SOFTWARE: Word Perfect 6.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/864,038A
 FILING DATE: May 28, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-184459
 FILING DATE: 15-July-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: C. Bruce Hamburg
 REGISTRATION NUMBER: 22,389
 REFERENCE/DOCKET NUMBER: F-5610
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212)986-2340
 TELEFAX: (212)953-7733
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 738
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Plinctada fucata
 CELL TYPE: mantle epithelial cell
 FEATURE:
 NAME/KEY: peptide
 LOCATION: from 1 to 738
 IDENTIFICATION METHOD: E (by experiment)

Query Match	40.0%;	Score 593.5;	DB 3	Length 738;
Best Local Similarity	41.1%;	Pred. No. 3.2e-41;		
Matches 167; Conservative	10;	Mismatches 82;	Indels 147;	Gaps 12;

```

QY      11  GSNAGSGRGGLGCGCAGCAAAAAAAAAAAGCAQ-----GCTGGGCGSOGTSG---54
Dh      113  GCMCAAGCGCGAGGAGGAGAGCAAGACAGAGAGLGLGLGGLGGLGGLGGLGGLGCD 1727
QY      55  -----RGLGGGCGCAAAAAAAAAAAGCAGCGCGYGGLGSGCTSGRGLGCG 99
Dh      173  DLFLDLPDLDLCAALATGACGAGAGCAAAAAAAAAAAGCGVGCAAAAAAAAAAGGAGR 2322
QY      100  --GAGAAAAAAAAAAGAGC-----QGGYGGGSGGTGSRGGLGCGC--ACAAAAAAAAAA 150
Dh      233  LGGCAAAAAAAAAAAGAGAGGLGGLGGLGGLGGLGGLGGLGGLGGLGGYGGSAAAAAAAAA 2911
QY      151  GGAAGGGYGGGSGGTGSRGGLGGLGCGAG-----AAAAAAAAAAGCAGCGYGGGLGSGC 2033
Dh      292  AAMGGGLGGYGTG--GRCGRRCRGRGRRRAAAAAAAAAAAGGGGGGGGG-----3444
QY      204  TSGRGLGCGCAGCAAAAAAAAAA-----226
Dh      345  ----GGGGGAGAAAAAAAAASASASRQMSGIRDALGDIRDLRSNCAKASAKASAVA 4000
QY      227  -----GCAAGCGYGGGLGSGGTSGR--245

```

—
—
—
—
—
—
—
••
—

Db 401 STKQIDDLKVLVDLNLGSLKSSASASASASASAGCGGGCGGCGGGCGGGGAGAL 460

0Y 246 -----GGLGGGAGAAAAAAAAGGAGGGGCGGLSGQTSG 282

Db 461 AALAAAGAGGGLGCGCGGALAAALAA--GAGGGGCGFGLGGLG 505

RESULT 7

```

US-08-529-190B-1
: Sequence 1, Application US/08529190B
: Patent No. 5833991
:
: GENERAL INFORMATION:
: APPLICANT: Masucci, Maria G
: TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES
: TITLE OF INVENTION: CONFERRING INVISIBILITY TO THE IMMUNE SYSTEM
: NUMBER OF SEQUENCES: 76
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner & Wilcoff, Ltd.
: STREET: One Financial Center
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02111
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: Wordperfect 6.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/529,190B
: FILING DATE: 15-SEP-1995
: CLASSIFICATION: 514
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: SE9501324-9
: FILING DATE: 10-APR-1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US08/522,595
: FILING DATE: 01-SEP-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Williams, Ph.D., Kathleen A
: REGISTRATION NUMBER: 34,360
: REFERENCE/DOCKET NUMBER: 3255/53015
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-345-9100
: TELEFAX: 617-345-9111
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 235 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: unknown
:
: MOLECULE TYPE: protein
:
: US-08-529-190B-1

```

Query Match	37.88;	Score 561;	DB 2;	Length 235;
Best Local Similarity	51.28;	Pred. NO. 5e-39;		
Matches 133; Conservative	5;	Mismatches 92;	Indels 30;	Gaps 9

[illegible]

Db 164 -----AGAGGAGAGAGAGGAGGAGA-----GGGAGGAGAGGCGAGAGAGGAGGAGGAGGAGA 214

QY 260 AAAAAAGAGGCGGyGGGSGG 279

Db 215 GAGGAGGAGAGGAGGAGAGGAG 234

RESULT 8

US-08-425-069-4
Sequence 4, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Hlman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-4

Query Match	37.3%;	Score 554;	DB 1;	Length 595;
Best Local Similarity	44.1%;	Pred. No. 4.1e-38;		
Matches 166;	Conservative 15;	Mismatches 101;	Indels 94;	Gaps 17;

[illegible]

Db 325 GGYGP-GGQGGPGYGPQQGPGYGPSSAAAAAAGPQQGPGYGP-GGQGGPGPSA 382
QY 247 -----GLGGGAG-----AAAAAAAAAG-----GAGGGYGG 274
Db 383 SAAAAAAGPBGYPGQQGPGYGPQQGPGSPGSASAAAAAAGPBGYPGQQGPGG 442
QY 275 L--GSQGTSGIRRPAA 288
Db 443 YAPGQGGPGSPGSAAA 458

RESULT 9

US-08-317-844B-4
; Sequence 4, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-317-844B-4

Query Match 37.3%; Score 554; DB 2; Length 595;
Best Local Similarity 44.1%; Pred. No. 4,1e-38;

Matches 166; Conservative 15; Mismatches 101; Indels 94; Gaps 17;

QY 3 GSHNNHNSMASRGGLGGGAGAAAAAAGAGAGQ-----GGYG-----GLGSO 50
Db 87 GGYGPQQGPGYGPQQGPGSPGSAAAAAASAESGQQGPGYGPQQGPGYGPQQ 146
QY 51 GYSGRG-----GLGGGAGAAAAAAGAGAGCGYGGYGSQTSGRG-----GLGGGAG 102
Db 147 GPGYGPQQGPGSPGSAAAAAAGPQQGPGYGP-GQQGPGYGPQQGPGSPGSA 205
QY 103 AAAAAAAGAGAGCGYGGYGSQTSGRG-----GLGGGAGAAAA-----144
Db 206 AAAAAAGSGGQGGPGYGP-GQQGPGYGPQQGPGSPGSAAAAAAGPQQGPGYGP 264

QY 145 -----AAAAAGAGCGGYG-GLGSGQTSGRGLGGGAGAAAAAAGAGAGQ 193
Db 265 GGQGGPGSPGSAAAAAAGPBGYPGQQGPGYGPQQGPGSAGSAAAAAAGGQGG 324
QY 194 GYGGLGSGQTSGRG-----GLGGGAGAAAAAAGAGAGCGYGGYGSQTSGRG-- 246
Db 325 GGYGP-GGQGGPGYGPQQGPGYGPSSAAAAAAGCGQGGPGYGP-GGQGGPGPSA 382
QY 247 -----GLGGGAG-----AAAAAAAAAG-----GAGGGYGG 274
Db 383 SAAAAAAGPBGYPGQQGPGYGPQQGPGSPGSASAAAAAAGPBGYPGQQGPGG 442
QY 275 L--GSQGTSGIRRPAA 288
Db 443 YAPGQGGPGSPGSAAA 458

RESULT 10

US-07-609-716-31
; Sequence 31, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappelletto, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-609-716-31

Query Match 35.0%; Score 520; DB 1; Length 1177;
Best Local Similarity 43.9%; Pred. No. 4,1e-35;

Matches 122; Conservative 22; Mismatches 128; Indels 6; Gaps 4;

QY 11 GSWASGRGLGGGAGAAAAAAGAGAGCGYGGYGSQTSGRGLGGGAGAAAAA 70
Db 71 GSGAAGYAG 130
QY 71 AAGG-GAGQGGYGGYGSQTSGRGLGGGAGAAAAAAGAGAGCGYGGYGSQTS 129
Db 131 SGAAAGYAG 190
QY 130 GRGGLG---GGGAGAAAAAAGAGAGCGYGGYGSQTSGRGLGGGAGAAAAA 186

[illegible]

RESULT 11
US-08-175-155-29

```

Sequence 29, Application US/0815155
Patent No. 5641648
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W.
APPLICANT: Dorfman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
TITLE OF INVENTION: Repetitive DNA
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Honbach, Test, Albritton & Herbert
STREET: Four Embarradero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IS-08-175-155-29

```

Query Match	35.08	Score	520	DB 1	Length	1177
Best Local Similarity	43.98	Pred. NO.	4.1e-35			
Matches 122	Conservative	22	Mismatches	128	Indels	6
					Gaps	4

Qy	11	GSMASGRCLCGOCAGAAAAAAAACGACGGCGTGLGSGTSGRGLGGGCGAAAAAA	70
Db	71	GSAAAGTGCAGAGSACAGAGSCGAGAGSAGAGSAGAGSACAGAGSACAGAGSAGAG	130
Qy	71	AAAAG-GAGCGGTGTLGSGTSGRGLCGOCAGAAAAAAAAGGACGGGTGTLGSGGTS	129
Db	131	SGAAGTGCAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGS	190
Qy	130	GRRGLG---GGAGACAAAAAAGAGAGCGGTGTLGSGGTSGRGLGGGAGAAAAAAA	186
Db	191	GAAATGCAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG	250

[illegible]

RESULT 12
US-08-477-509B-64
; Sequence 64; Application US/08477509B

Patent No. 5/770697
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A
APPLICANT: Cappello, Joseph
APPLICANT: Crisman, John W
APPLICANT: Dorman, Mary A
TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,509B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecautin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-509B-64

[illegible]

```

0Y 71 AAAAG-CAGCGGTGGCLGSGTSGRGCLGGCGAGCAAAAAAAGAGAGCGGTGGCLGSGTGS 129
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 SGAAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 190
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 130 GGGGLG---GCGAGAAAAAAGAGAGCGGTGGCLGSGTSGRGCLGGCGAGCAAAAAA 186
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 GAACTGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 250
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 187 AAG-GAGCGGTGGCLGSGTSGRGCLGGCGAGCAAAAAAAGAGAGCGGTGGCLGSGTSGR 245
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 AAGGAGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 310
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
0Y 246 GGLG-GGAGCAAAAAAAGAGAGCGGTGGCLGSGTSG 282
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 AGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 348
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-707-237A-35
; Sequence 35, Application US/08707237A
; Patent No. 5830713
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Capello, Joseph
; APPLICANT: Crisman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
; TITLE OF INVENTION: REPETITIVE DNA
; NUMBER OF SEQUENCES: 108
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/707,237A
; FILING DATE: 03-SEP-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-10/WHM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 35:

```

```

SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-707-237A-35

Query Match      35.0%; Score 520; DB 2; Length 1177;
Best Local Similarity 43.9%; Pred. No. 4.1e-35;
Matches 122; Conservative 22; Mismatches 128; Indels 6; Gaps 4;

QY 11 GSMASGRGCLGCGAGAAAAAAAAGACGGCGYGGLSGDTSRGLGCGAGAAAAAA 70
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 GSGAGYGAAGAGSAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS 130
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 71 AAAAG-GAGOGGYGGLSGDTSRGLGCGAGAAAAAAAAGCAGOGGYGGLGSQGTG 129
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 131 SGAAHYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS 190
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 130 GRGGLG---GCGAGAAAAAAAAGCAGOGGYGGLGSQGTGSRGLGCGAGAAAAAAA 186
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 191 GAAGYGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS 250
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 187 AAG-GAGOGGYGGLGSQGTSGRGGLGCGAGAAAAAAAAGCAGOGGYGGLGSQGTSGR 245
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 251 AAGYGAAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA 310
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 246 GLG-GCGAGAAAAAAAAGCAGOGGYGGLGSQGTSG 282
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 311 AGYGAAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA 348

RESULT 14
US-08-482-085B-64
Sequence 64, Application US/08482085B
Patent No. 6018030
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Richardson, Charles
APPLICANT: Chambers, James
APPLICANT: Causey, Stuart
APPLICANT: Pollock, Thomas J.
APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W.
TITLE OF INVENTION: NO. 6018030el Peptides Comprising Repetitive
TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
NUMBER OF SEQUENCES: 112
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,085B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049

```

T31328
fibroin - Chinese oak silkwmoth
C:Species: Antheraea pernyi (Chinese oak silkwmoth)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukubiro, K.,
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, Anth.
A:Reference number: 220995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-2639 <SE2>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C:Genetics:
A:Introns: 14/3

Query Match	41.68%	Score 617.5;	DB 2;	Length 2639;
Best Local Similarity	39.88%	Pred. No. 8.8e-30;		
Matches 174;	Conservative 10;	Mismatches 90;	Indels 163;	Gaps 13

Oy	14	ASGRCGLG-----	OGAGAAAAAAAAAGAGCGGCGVGLGSGCT-----	52
Db	1142	ASGAGCGSGCGTGMWDGGTGSDSAAAAAAAAAAAAAAAAA	SGAGSGGCTGTGTSDSAAAAAAAAA	1201
Oy	53	-----SGRGLG-----	OGAGAAAAAAAAAGAGCGGCGVGLGSGCT-----	90
Db	1202	AAAAAGSGAGCGGCGVGMWDGGTGSDSAAAAAAAAA	SAAGSGGCGTGTGTSDSA	1261
Oy	91	-----SGRCGLG-----	QGAGAAAAAAAAAAGAGCGGCGVGLGSGCT-----	121
Db	1262	AAAAAAAAAAGSGAGGAGCGGCGVGMWDGGTGSDSAAAAAAAAA	AGAGCGGCGVGLGSGCT-----	1321
Oy	122	-----GLGSGCTSGRCGLG-----	GLGSGCTSGRCGLG-----	136
Db	1322	GSSAAAAAAAAAARRAGHDRAAGSAAAAAAAAA	SAAGSGGCGGCGVGMWDGGTGSDS	1381
Oy	137	-----QGAGAAAAAAAAAAGAGCGGCGVGLGSGCT-----	SGRCGLG-----	174
Db	1382	SAAAAAAAAAAASGAGSGGCGVGLGSGDSAAAAAAAAA	AGAGCGGCGVGLGSGCT-----	1441
Oy	175	-----OGAGAAAAAAAAAAGAGCGGCGVGLGSGCT-----	OGSGRCGLG-----	212
Db	1442	GYSGDSAAAAAAAAAAGAGSGGCGGCGVGLGSGDSAAAAAAAAA	SAAGCGGCGVGLGSGCT-----	1501
Oy	213	QGAGAAAAAAAAAAGAGCGGCGVGLGSGCTSGRCGLGCGAGAAAAAAAAA	AGAGCGGCGVGLGSGCT-----	271
Db	1502	YGSDSAAGAGAGCGGCGVGMWDGGTGSDSAAAAAAAAA	AGAGCGGCGVGLGSGCT-----	1561
Oy	272	YGLGSGCTSGIRPPA	288	
Db	1562	DGGYGS--GSSAAAAAAAA	1577	

RESULT 3
O0BE31
nuclear antigen EBNA1 - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 22-Oct-1999
C:Accession: C43043; S42440; A03773; S33021
R:Bankier, A.T.; Delinger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713
A:Accession: C43043
A:Molecule type: DNA
A:Residues: 1-641 <BAN>
A:Cross-references: EMBL:V01555; NTD:959074; PIDN:CAA24816.1; PID:g1334880
A:Experimental source: strain B95-8
R:Bankier, R.; Bankier, A.T.; Biggin, M.D.; Delinger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nucleic 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; MUID:84270667
A:Contents: annotation: protein coding region
R:Sample: J.; Hummel, M.; Braun, D.; Birkenbach, M.; Kieff, E.
Proc. Natl. Acad. Sci. U.S.A. 83, 5096-5100, 1986
A:Title: Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear proteins.
A:Reference number: S42440; MUID:86259739
A:Accession: S42440
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-66 <SAM>
A:Cross-references: EMBL: M13941; NID: g930399; PINN: AAA45889.1; PID: g955157
C:Superfamily: Epstein-Barr virus nuclear antigen
C:Keywords: DNA binding; transcription regulation

Query Match	41.1%;	Score 610.5;	DB 1;	Length 641;
Best Local Similarity	-48.6%;	Pred. No. 8.2e-30;		
Matches 143;	Conservative 14;	Mismatches 104;	Indels 33;	Gaps 10;

Oy	2	RGSHHHHHGSMAGRGGLG--GGAAGAAAAAAAAGAGAG--GGYGLISGQGTSGRGLG	59
Oy	3		64
Db	83	KGTH-----GGTGAAGAGAGAGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGAGGGA	130
Oy	60	GCGAGAAAAAAAAGAGAG--GGYGTGCSGCTSGRGLGCGAGAAAAAAAAG--GA	115
Db	131	GAGGAGAGAGAGAGAGGCGAGAGGCGAGGAGAG--GAGAGAGAGAGGCGAGCGAGAGAG	188
Oy	116	GGGGYGGLSGQGTSGRGLGGO--GAGAAAAAAAAGAGAGCGGCGYGLISGQGTSGRGLG	174
Db	189	GAGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGAGAGGAGAGAGAGAGG-----AGGAGAGG	244
Oy	175	QGAGAAAAAAAAGAGAGGAGGCGYGLISGQGTSG-----RGGLGCGAGAAAAAAAAGAGAG	230
Db	245	AGAGAGAGAGAGAGAGAGAGAGAGAGAGGCGAGCGAGCGAGAGAGAGAGAGAGAGAGAG	304
Oy	231	QGGYGTGCSGCTSGRGLGCGAGAGAAAAAAAAGAGAGCGGCGYGLISGQGTSGR	284
Db	305	AGGAGAGAGAGAGAGAGAGAGG-----RCRGCGGCGAGAGAGGCGGCGRGCGGCGRR	354

RESULT 4
 F70806
 hypothetical glycine-rich protein RV3508 - *Mycobacterium tuberculosis* (strain H37RV)
 C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70806
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garler, T.; Churcher, C.; Harris, D.; Gordon
 R.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajadram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70806
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1901 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAA17745.1; PID:g2922
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3508
 C:Superfamily: collagen alpha 1(IV) chain

[illegible]

Db 476 AGAGGAGGAAGTGGTGGVYVGAAGKAGTGGTGAGGAGGASGATATATGATGCTGCTGCG 535

Qy 120 YGGL-----GSGCTSGRGLGGGAGAGAAAAAAAGAGAG-----GGYGLGSGCTGSR 169

Db 536 AGGAGGAGAGNTGGTGTGNSGCGG--GTGGAGGAGAGAGVGVADNPCTIGGTGGTGGG--GA 592

Qy 170 GGLGCGG--AGAAAAAAAGAG--AGGCGYGLGSGCT--SGRGLGGGCGAGAAAAAA 223

Db 593 GGAGCGGGSSSAGAGTNGSGAGCTGGCGGAGAGAGAGADNPFGTIGAGAGTGTGGAAG 652

Qy 224 AAAGAGGCGGTYG-----LGSCTSGRGLGGG--GAGAAAAAAAGAG-----GGG 271

Db 653 GAGGATGCTGTGGAGATGSGTGGTGGTGGAGAGAGAAAAAGSSATGCGAFAGAGC 712

Qy 272 YGGLGSGCTSGT 283

Db 713 ECGAG--GNSGV 722

RESULT 5
 D70807
 hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: D70807
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fieldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, R.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: D70807
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1489 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PID:CAAL17751.1; PID:g2924445
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3514
 A:Superfamily: collagen alpha 1(IV) chain

Query Match	37.4%	Score	556	DB	2	Length	1489
Best Local Similarity	39.3%	Pred. No.	2.5e-26				
Matches	150	Conservative	17	Mismatches	103	Indels	112
						Gaps	16

Oy	11	GSMAGRGGLGGO--GAGAAAAAAAAAGGAGCGGCGVGLGSQ-----GTS	53
Db	292	GGALGCTGCTGTGGAGGAGCGRALLLGAAGCGGGLGAGAGGCGTGCAGCDGVLGSGVTG	351
Oy	54	GRGGLG--GAGAAAAAAAAAAAAAGAGCGCGGGLGSGQTSGRGGLGQCA-----	101
Db	352	GKGGCGVAGLGGAGAGCAAGOLFASAGAGMAGVSGAGCG--GDSGAGAGAGADADPGAT	409
Oy	102	-----GAAAAAAAAAG-----AGGGGGLGSGQT--SGRGLGAGGAGA	141
Db	410	GCTGFAGGAGAGGAGCGSSGAGGTNGSGAGCGGAGGAGAGADNPITGIGTGDGGTG	469
Oy	142	AAAAAAAAAGGAGCGCGYGC-LGSOQTSGRGGLGQ-----GAGAAA-----A	182
Db	470	GAAAGGAGGAAGTCTGTMGTTCMAGVGGAGGGGAGGAGADADDPGATGCTGFA	529
Oy	183	AAAAAAGA-----GOGGYGGLGSQ-----TSGRGLGAGGAGAAA	220
Db	530	GGAGGAGGAGSSGAGGTNGSGGAGCTGGGGAGGAGGAGADNPITGIGTGGDGTGCA	589
Oy	221	AAAAAAAAAGGAGCGYGC-LGSOQTSGRGGLGQ-----GAGAAAAAA--AAA	264
Db	590	GAGGAGGAAGTGTGTMGTTCMAGVGGAGGCGDGGAGAGADADDPGATGCTGFA	649
Oy	265	GGAGCG-----YGLGSGQCTSG	282

Db 650 GGAGKAGSSSAGGTNSSGSAG 671

RESULT 6
A44112
spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)

C:Species: Nephila clavipes
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998
C:Accession: A44112: S27624
R:Hitman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A:Title: Isolation of a cDNA encoding a second dragline silk fibroin. Nephila clavipes
#:Reference number: A44112; MUID:92406876

A:Accession: A44112
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-627 <H1N>
A:Cross-references: GB:M92913; NID:g159713; PTD:g159714
A:Note: sequence extracted from NCBI backbone (NCBIP:111893)
R:Hitman, M.B.; Lewis, R.V.
submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephlila
A:Reference number: S37824
A:Accession: S27824
A:Molecule type: mRNA
A:Residues: 19-627 <H12>
A:Cross-references: EMBL:M92913

Query Match	37.3%	Score 554	DB 2	Length 627
Best Local Similarity	44.1%	Pred. No. 1.7e-26		
Matches 166	Conservative	15	Mismatches 101	Indels 94
				Gaps 17

```

OY      3 GSHHHHHHSSAAGCGGGLGGCGACGAAAAAAAAAAAGCAQ---GGYG-----GLGSQ   50
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      87 GGYGP RQQGPGG YGGGQGPGSPGSAASASAESAESQDQGPGGYGPQQGPGGYGPQQ    146
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      51 GTSGRG----GLGGGAGCAAAAAAAAAGAGAGGCGGGLGSOGTSGRG----GLGGCAG   102
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      147 GPGGCGPQQGPGSGGSAASAAAAASGPGQQGPGGYGP-GQQGPGGTGPGQQGPSGPSA   205
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      103 AAAAAAAAAGAGAGCGGCGGLGSOGTSGRG----GLGGCAGAAAA-----144
          ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      206 AAAAAAASGPGQQGPGGYGP-GQQGPGGTGPGQQGLSLSPSAAAAAAMPQQGPGGYGP   264
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      145 -----AAAAAAGAGAGCGGYG-GLGSOGTSGRGLGSOGAGAAAAAAGAAGQ   193
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      265 GQQGPGSPGSAASAAAAAAGPGYGPQQGPGGYGPQQGPGSAGSAAAAAAPQQGL   324
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      194 GGYGGLGSOGTSGRG----GLGGCAGAAAAAAA--GAGCGGGLGSOGTSGRG--   246
          ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      325 GGYGP-GQQGPGGYGPQQGPGGYGPGSASAAAAAGCGQQGPGGYGP-GQQGPGSPSA   382
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      247 -----GLGGCAG-----AAAAAAG--GAOCGGCG   274
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      383 SAAAAAAGAPGGYGPQQGPGGYAPGQQGPSGPGSASAAAAAAGPGYGPQQGPGG   442
          | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY      275 L--GSOGTSGIRRPAA   288
          | | | | | | | |
Db      443 YAPGQQGPSGPGSAAA   458

RESULT       7
E70806
hypothetical glycine-rich protein RV3507 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70806
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, B.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Kjandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
```

Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Accession number: A70500; MUID:98295987
A:Accession: E70806
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1381 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17744.1; PID:g29244444
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3507
A:Superfamily: collagen alpha 1(IV) chain

	Score	DB 2, Pred. No. 1.1e-25, 17; Mismatches 115;	Length 1381; Indels 105; Gaps 13;
Query Match	36.7%		
Best Local Similarly	37.8%		
Matches 144; Conservative			

QY 1 MRGHHHHHNSMASRGCGGAGAGAAAAAAGCAGCGGCGL--GSQGTSGRGL 58
| | : : : | | | | | : | | | : | | | :
Db 205 MCGTCGNGCNCALLLGGGGL--GAGCMGTCGCGTCGTCGNCNCALLLGAAGVGAGGI 262

```
Oy 59 GGCGAGAAAAAAGAGGCGGGYGTGSGTGTGRGGLTCGAGAAAAAAGCA --- 115
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 263 GGGCTGAGAGAAGAGCTG - - GNGAGAGLEMG - - GDGAGGCGGDGAAGDAASAAGTGK 318
```

```

QY 116 -----GGGCGGCGAGGCGTSGRGGAGGCG-----AG 140
      | | | | | | | | | |
DB 319 GGGCGDGGTGGAGGAGPVLFGHGAGAGGAGGCGGCGMCGAGGDDTTVIAAGTGGGCGTGG 378

```

Oy 141 AAAAAAAAAAGGAGCGG--YGLGSQGTSGRGILGGGCGAIAAAAAAAAG---GAGGG 194
| | | | | : | | | | | : |
Db 379 AAGAGCAAGARGALTSSGLAGVGAGTGGTGCGNGADAAAIVVFGANGDPGFAGGKG 438

```

Qy 195 GYGGLGSG-----GTSRGGGLGQ-----GAGAAAA 221
      | | | | |
Db 439 GNGGTGGAATGTVAGADGCTGGKGGTGAGAGACNDAGSTGCPGKGGGDDGCTGGAGAGAGA 498

```

```

Qy      222  AAAA---AGCAGCGCGTGLSGSGTSGAGCGGCG-----GAGAAAAAA 261
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      499  AGTGGCGHAGTGGDGGGTGGNGGNTGGVNGADNTLPDTPCGAGGPGCAGGAGG 558

```

```

QY      262 AAAGCAGGCGTGGGSGTSG 282
          || | | | | | | : |
DB      559 AAGCGGTGCTGGNGGNGG 579

```

RESULT 8
B70807
[unclear] studies with methods. [unclear] Microbacterium tuberculosis (strain H37Rv)

hypothetical glycine-rich protein RV3512 - *Mycobacterium tuberculosis* (str
C:Species: *Mycobacterium tuberculosis*
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

R.Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon, R.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Geetha, S.; Hamilton, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

A: Authors: Sgares, R., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G.
A: Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A: Reference number: A70500; MUII:98295987

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1079 <COL>

A:Experimental source: Strain H3/Rv
C:Genetics:
A:Gene: RV3512

Query Match 36.6%; Score 543.5; DB 2; Length 1079;
C:supernatantly; collagen alpha 1(+), chain; fibrillar collagen catalo; terminal homology

Best Local Similarity 40.7%; Pred. No. 1.1e-25;
Matches 144; Conservative 12; Mismatches 109; Indels 89; Gaps 14;

QY 14 ASGRGRLGCGGA-GAAAAAAAAAAGCGGQGGYCGLSQGTG-----GRG 56
| : | | | | | | : | | | | | | | : | :
Db 217 ANGLAGNGGCGDGAAGAVGTSATGAGDGGHGCTGAACGNGCGTCGAGGSSIDVGCGTG 276

```

QY      57  GLGQGAGAAAAAAGAG-QGCGGLGSG-----GTGGRGLGQG- 100
          | | | | | | | | | | | | | | | | | | | | | |
Db      277 GTGCGNGNGAIGGAGDAGGSSNGSGGIGGACGNAGGAAGSNGGTGANGTGGDG 336

```

```
QY      101 -AGAAAAAAGGAGGGGCGGSGGTGCGGCG-----AGAAAAAA 148
          ||| | : |||| | | | |||| | |
Db      337 NGGAGAATAGSNGCAGTGSACGNG--CTGRGCGSGCAGCDIGCGCKGCGNGACGEVG 394
```

```
QY      149 AAGGAG-----QGCGTGLGSQGTSSRGGTGGGAGAAAAAAGAAGGCGGYCL--   199
          |||||           || || | || | | | | || | | || || |
Db      305 GAGGAGSGPNTSPGGNGGGGGGGSSGAG-GAAAGAGAGGGGANCTAGNNGGGAGGTGG   453
```

```

QY 200 -----GSGCTSGRGGLGCGAGAAAAAAGGACGGCGGCGTSGRG 246
      | : | | | | | | | | | | | | | | | | | | | | | |
Db 454 AGAASATNGSGGACGTCGDDGSGGACGTGCAGCTGCAAGDCGGCGGAGG- GAGGG 512

```

```

QY 247 GUGGG-----ACAAAAAAAAAAGGA-GGGY-GGLSSQGTSG 282
      |||      ||| ||| ||| ||| ||| ||| :|
Db 513 GAGGAGGTGNGNITGCTAGTAGAAGNGAGKCGAGCGGGTGGTGGGGAG 566

```

RESULT 9
H70846

C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; H
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70846

A: Molecule type: DNA
A: Residues: 1-1538 <COL>
A: Cross-references: GB:AL021841, GB:AL123456, NID:g3261517, PIDN:CAAI7117.1, PID:g326

C:Genetics:
A:Gene: RV345C
C:Superfamily: collagen alpha 1(IV) chain

Query Match 36.28; Score 538; DB 2; Length 1538;

QY	11	GSMASGRGGLGCGAGAAAAA	GGAGGCGGCGLGSQCT	5GRRGLGCGGAGAAAAA	70
Matches	126;	Conservative	13;	Mismatches	131;
				Indels	54;
				Gaps	

Db 1030 GMINGGLGFGCGAGCGGAVDVAATTGGAGCGAGGFASTGLGPGGAGGPGGAGDFAAG 1088

QY 71 AAAAGG-----GGGGGGLGSG-----TSGRGLGGGGAGAAAAAAGG----- 114

Db 1090 VGGVGGAGGCGGAGGVGGFGGCGGIGGEERTGTGNGSGGDDGGGGISLGNGGLGNCGVS 114
QY 115 -----AGCGGYGGLSQGTSGRGLGCGGAGAAAAAAGAAGAGGGYGGLSQ--- 164

Db 1150 ETGEGGAGGNGYGPG--GPEGNGLGCGNGAGGNGVSTTGCDGAGGKGGNGGCGGN 1200

QY 165 -----GTCGRGGLGCGGAGCAAAAAAAGC-----AGGGYGGLSGGTS 2050

Db 1208 VGLGSDAGSGAGGNGCIGTDACGAGGAGGAGCGSSKSTTTGAGSGGAGGNGGTGLN 1209
QY 206 GRCGTCGGGAGAAAAA--AAAGGACGGGTCGTCGTS-GRCGLGGGGAGAAAAA 261

QY	Db	QY	Db
170	GC--TGGCGAGAAAAAAGAGAGCGGCGYGGJGSGOSTSGGCGTGGGAGAAAAA	225	
232	GGGFCAGGAGGAGGCGGCGGCGGCGGGLGGGSHGGGFCAGGCGTGGGAAAGV	291	
226	AGGAGCGGCTGGLGSGGCTGGRGLGCGGAGAAAAAAGAGCGGCGYGLG	276	
292	GGGGGFGGGGGGCGGSGHGGGFCAGGAGGCGGLGCGGGAGAGGGGTLG	342	

RESULT 13
A70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium mageritensis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: A70896
R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1 853 <COL>
A:Cross-references: GB:AL021897; GB:AL123456; NID:q3256022; PIDN:CAAL17207.1; PID:el2519
A:Experimental source: Strain H37Rv
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match	35.0%;	Score 519.5;	DB 2;	Length 853;
Best Local Similarity	41.78;	Pred. No. 2.4e-24;		
Matches 134; Conservative	14;	Mismatches 122;	Indels 51;	Gaps 13

QY 3 GSHHHHHHSMASGRGGLGQ_GACAAAAAAAAAAAAAGAGOGGYGGLSGQ--TSRGRGLG 59
Db 293 GSGGH-----ALLMAGGAGAGNGSGGTGCAGGCTTAGAGNGAGAGGGGGTGLLFFGCGAG 348
QY 60 CGGA-----CAAAAAAAAAAGAGAGOGGYGGLSGQGTSGRGGLGGQ-----AGAAAAAAAA 110
Db 349 GGCAGAGNLLAAGNGVSSSGGGGAGGTGAAGDG--GAGGAGGNRLMVGAGAGGAGG 406
QY 111 AAGGAGOGGYGGL-----GSGTSGRGGLAGGAGAAAAAAAAAAAAAGAGOGGYGGLSGQ 164
Db 407 GAGGAGGKGGSGSLSGMANGAGGDSRGGTG--GAGGEGGAAGLVTGTGGHGGDSAG-- 462
QY 165 GTSGRGGLGGOGGAAAAAAAAAAAAAGAGAGOGGYGGLSGQTS-----GRGLGGGAGAGA 217
Db 463 GAAYKCGDGAAGATGTAGAGRGAGSGSGGCGDGGGAGPAGLFFDGGAGCGNGGAA 522
QY 218 AAAAAAAAAAG-----AGGGYGGGLSGQTS-----GRGLGGGAGAAAAAAAAAAAA 262
Db 523 AAGGAGGGAGGCGGNGNGNGNGNGNGATGTGLYNGGAGGGGAATAGAGGAGANGV 582
QY 263 -AAGGAGOGGYGGLSGQTS 282
Db 583 SSTNGGCTGANGCIGTGGSG 603

RESULT 14
F70963
hypothetical glycine-rich protein RV2634c - *Mycobacterium tuberculosis* (strain H37RV)
C.Species: *Mycobacterium tuberculosis*
C.Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C.Accession: F70963
R.Cole, S.T.: Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Comor, R.; Bross, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holtrold, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70963
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <COL>
A:Cross-references: GB:D80225; GB:AL123456; NID:g3342265; PIDN:CAB02341.1; PID:e26639
A:Experimental source: Strain H37Rv

A;Gene: RV2634C
C;Superfamily: unassigned collagens

Query Match	Score	DB 2;	Length
34.88;	516.5;	DB 2;	778;

Best Local Similarity 39.4%; Pred. No. 3.3e-24;
Matches 134; Conservative 14; Mismatches 121; Indels 71; Gaps 15;

QY 10 HGSMSGRGG-LGGGAGAGAAVAAAAGSAGQQGIGGL-GSGGIGSNGBDDBGGCNCN..VV
:::| | | | | | | | | | : | | | | |

Db 393 NGTDNSGNGNQTGGNGGPGPAGVGGEAGVGGGGGLGESLDGNDGTGKGKGAGGTAGTDG 452

67 AAAAAAAAAAGGAGC-----GGYGGLGSQGT-----G 92

1 | | | : ||| | :
 2 | | | : ||| | :
 3 | | | : ||| | :
 4 | | | : ||| | :
 5 | | | : ||| | :
 6 | | | : ||| | :
 7 | | | : ||| | :
 8 | | | : ||| | :
 9 | | | : ||| | :
 10 | | | : ||| | :
 11 | | | : ||| | :
 12 | | | : ||| | :
 13 | | | : ||| | :
 14 | | | : ||| | :
 15 | | | : ||| | :
 16 | | | : ||| | :
 17 | | | : ||| | :
 18 | | | : ||| | :
 19 | | | : ||| | :
 20 | | | : ||| | :
 21 | | | : ||| | :
 22 | | | : ||| | :
 23 | | | : ||| | :
 24 | | | : ||| | :
 25 | | | : ||| | :
 26 | | | : ||| | :
 27 | | | : ||| | :
 28 | | | : ||| | :
 29 | | | : ||| | :
 30 | | | : ||| | :
 31 | | | : ||| | :
 32 | | | : ||| | :
 33 | | | : ||| | :
 34 | | | : ||| | :
 35 | | | : ||| | :
 36 | | | : ||| | :
 37 | | | : ||| | :
 38 | | | : ||| | :
 39 | | | : ||| | :
 40 | | | : ||| | :
 41 | | | : ||| | :
 42 | | | : ||| | :
 43 | | | : ||| | :
 44 | | | : ||| | :
 45 | | | : ||| | :
 46 | | | : ||| | :
 47 | | | : ||| | :
 48 | | | : ||| | :
 49 | | | : ||| | :
 50 | | | : ||| | :
 51 | | | : ||| | :
 52 | | | : ||| | :
 53 | | | : ||| | :
 54 | | | : ||| | :
 55 | | | : ||| | :
 56 | | | : ||| | :
 57 | | | : ||| | :
 58 | | | : ||| | :
 59 | | | : ||| | :
 60 | | | : ||| | :
 61 | | | : ||| | :
 62 | | | : ||| | :
 63 | | | : ||| | :
 64 | | | : ||| | :
 65 | | | : ||| | :
 66 | | | : ||| | :
 67 | | | : ||| | :
 68 | | | : ||| | :
 69 | | | : ||| | :
 70 | | | : ||| | :
 71 | | | : ||| | :
 72 | | | : ||| | :
 73 | | | : ||| | :
 74 | | | : ||| | :
 75 | | | : ||| | :
 76 | | | : ||| | :
 77 | | | : ||| | :
 78 | | | : ||| | :
 79 | | | : ||| | :
 80 | | | : ||| | :
 81 | | | : ||| | :
 82 | | | : ||| | :
 83 | | | : ||| | :
 84 | | | : ||| | :
 85 | | | : ||| | :
 86 | | | : ||| | :
 87 | | | : ||| | :
 88 | | | : ||| | :
 89 | | | : ||| | :
 90 | | | : ||| | :
 91 | | | : ||| | :
 92 | | | : ||| | :
 93 | | | : ||| | :
 94 | | | : ||| | :
 95 | | | : ||| | :
 96 | | | : ||| | :
 97 | | | : ||| | :
 98 | | | : ||| | :
 99 | | | : ||| | :
 100 | | | : ||| | :

[illegible]

QY 93 RGGGCGAGAAAAA--AAAAGGAGGGGG--LGSQGTSGKGGGGGCGAGAAA--AAA 140

Db 513 DGGICDNGALGAGGNGGTGGAGGNGRGGMLIGNGAGGAGGTGTTGGGAGGAGG 572

147 AAAGGAG-----OGGYGLGS-OGTSGRGLGGGA----GAAAAAAAAAAGGAGG 193

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Dd 3 / 3 VGGAGGEGELIDGAHIAEEBIBGBDBDGDVGBVVGBIUGNIGUDDDUVCGNGG:N:GGB=-----

QY 194 GGYGGLSQGTSGRGGLGQ---GAGAAAAAAGGAGGCGYGGGGSQGTSGRGGLG 243

Db 633 GGVGGTG--GIGGIGGAGGNGGAGGAGTTGGGATICGGGGTGGVGGAG--GTGGTGAG 688

250 GG--GAGAAA-----AAAAAAGGAGGOGGYGGTGSOGTSG 282

[illegible]

Db 689 GTTGGSSGAGGLIGWAGAAAGGIGAGGIGGGGAGGGGGG / 20

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843

RESULTS 15
A70934

hypothetical glycine-rich protein Rv0578c - *Mycobacterium tuberculosis* (Str. H37Rv)

C; Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C/Accession: A70934
R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris,

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, I.

Rajandream, M.A.; Rogers, J.; Kulter, S.; Seeger, A.; Breckon, D.; Ogunade Nature 393, 537-544, 1998

A; Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, A.; and others. Describing the biology of *Mycobacterium tuberculosis* from the genome.

A;Reference number: A70500; MUID:98295987

A;Accession: A70934
A:Status: preliminary: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1306 <COL>
A;Cross-references: GB:AL021942; GB:AL123456; NID:q3242298; PIDN:CAA17449

A;Experimental source: strain H37Rv

C:Genetics;
A:Gene: RV0578c

C;Superfamily: collagen alpha 1(IV) chain

Query Match	34.68;	Score 514;	DB 2;	Length 1306;
Post Local Similarity	42.08;	Pred. NO.	6.8e-24;	

Matches 131; Conservative 20; Mismatches 125; Indels 36; Gaps

3 GSHHHHHGSMASGRGGLGCGA-----AGCGGYGGLGSD---- 50

[illegible]

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:20 ; Search time 37.35 Seconds
(without alignments)

266,890 Million cell updates/sec

Title: US-09-490-291-2

Perfect score: 1485
Sequence: 1 MRGSHHHHSGMSASGRGL.....YGGLSGCTSGTRPANKLN 291

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1015	68.4	747	1	SPDL_NEPCL
2	610.5	41.1	641	1	EBN1_EBV
3	556.5	37.5	1901	1	YZ08_MYCTU
4	554	37.3	627	1	SPD2_NEPCL
5	525	35.4	384	1	GRP1_PERHY
6	516.5	34.8	778	1	YQ34_MYCTU
7	508.5	34.2	5263	1	FBOH_BOMMO
8	496.5	33.4	465	1	GRP2_PHAUV
9	496	33.4	603	1	YD25_MYCTU
10	494	33.3	914	1	WA22_MYCTU
11	490	33.0	338	1	GRP_ARATH
12	490	33.0	801	1	Y747_MYCTU
13	489	32.9	488	1	Y118_MYCTU
14	472	31.8	860	1	ELS_MOUSE
15	467	31.4	957	1	Y278_MYCTU
16	463	31.2	543	1	YP91_MYCTU
17	453.5	30.5	864	1	ELS_RAT
18	446.5	30.1	434	1	YK68_MYCTU
19	444.5	29.9	463	1	YK68_MYCTU
20	438	29.5	222	1	GRP1_PHAUV
21	418.5	28.2	515	1	Y140_MYCTU
22	409.5	27.6	481	1	LORI_MOUSE
23	394.5	26.6	747	1	ELS_BOVIN
24	385.5	26.0	750	1	ELS_CHICK
25	378	25.5	316	1	LORI_HUMAN
26	374	25.2	730	1	ELS_HUMAN
27	363	24.4	375	1	SANT_PLAFV
28	357.5	24.1	2038	1	FSH_DROME
29	341.5	23.0	401	1	CSP_PLACG
30	332.5	22.4	1733	1	VNUA_PRYKA
31	330.5	22.3	1380	1	DDX9_MOUSE
32	328	22.1	398	1	CSP_PLACC
33	311	20.9	1147	1	MYSB_ACACA

34	305	20.5	419	1	CSP_PLACM	P08676 plasmodium
35	304.5	20.5	183	1	GRP2_ORISA	P29834 oryza sativ
36	297.5	20.0	385	1	RO32_XENLA	P51992 xenopus lae
37	292.5	19.7	1373	1	CA21_MOUSE	Q01149 mus musculu
38	290	19.5	1027	1	CAFE_RIPPA	P30754 riftia pach
39	288.5	19.4	1366	1	CA21_CANFA	O46392 canis fami
40	284	19.1	1156	1	GLH4_CAREL	O76743 caenorhabd
41	283.5	19.1	404	1	CAZ_DROME	Q27294 dtrosophila
42	283.5	19.1	592	1	RB56_HUMAN	Q92804 homo sapien
43	283.5	19.1	622	1	K1C1_HUMAN	P35527 homo sapien
44	282	19.0	1366	1	CA21_HUMAN	P08123 homo sapien
45	281	18.9	672	1	PHX5_MOUSE	P08399 mus musculu

ALIGNMENTS

RESULT	ID	SPDL_NEPCL	STANDARD	PRT	747 AA.
1	ID	SPDL_NEPCL	STANDARD	PRT	747 AA.
AC	P19837				
DT	01-FEB-1991 (Rel. 17, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).				
OS	Nephila clavipes (orb spider).				
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;				
OC	Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.				
OX	NCBI_Taxid=6915;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=90384959; Pubmed=2402494;				
RA	Xu M., Lewis R.V.;				
RT	"Structure of a proteain superfiber: spider dragline silk.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).				
RN	[2]				
RP	SEQUENCE OF 653-747 FROM N.A.				
RX	MEDLINE=94165058; Pubmed=8120021;				
RA	Beckwith R., Arcidiacono S.;				
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentarius (Araneidae).";				
RL	J. Biol. Chem. 269:6661-6663(1994).				
CC	-1- FUNCTION: THIS SPIDER MAJOR AMPULLATE SILK POSSESSES UNIQUE CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF PEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED WITH ELASTIC AMORPHOUS SEGMENTS.				
CC	-1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.				
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; M37137; AAA29380.1; -				
DR	EMBL; U03848; AAB60212.1; -				
DR	PIR; A36068; A36068.				
KW	Silk; Repeat.				
FT	NON TER	1			
FT	DOMAIN	1			
FT	REPEAT	1	655		25 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	1	25		
FT	REPEAT	26	38		1.
FT	REPEAT	39	66		2.
FT	REPEAT	67	96		3.
FT	REPEAT	97	130		4.
FT	REPEAT	131	158		5.
FT	REPEAT	159	191		6.
FT	REPEAT	192	204		7.
FT	REPEAT	205	235		8.
FT	REPEAT				9.

```

FT REPEAT 236 262 10.
FT REPEAT 263 292 11.
FT REPEAT 293 305 12.
FT REPEAT 306 333 13.
FT REPEAT 334 360 14.
FT REPEAT 361 394 15.
FT REPEAT 395 424 16.
FT REPEAT 425 458 17.
FT REPEAT 459 485 18.
FT REPEAT 486 512 19.
FT REPEAT 513 525 20.
FT REPEAT 526 535 21.
FT REPEAT 536 556 22.
FT REPEAT 558 612 23.
FT REPEAT 613 642 24.
FT REPEAT 643 655 25.
FT REPEAT 662 662
FT CONFLICT 672 672
FT CONFLICT 672 672
FT CONFLICT 695 747

```

SEQUENCE 747 AA; 60528 MW; 850E44AB0D649E012 CRC64;

Query Match 68.4%; Score 1015; DB 1; Length 747;
 Best Local Similarity 59.2%; Pred. No. 3,8e-43;
 Matches 238; Conservative 8; Mismatches 14; Indels 142; Gaps 13;

```

QY 11 GSASGSGGGLGGGAGAAAAA-----AAAAAGAGGGGGY 45
DB 93 GSGAGAGGGLGGGAGAAAAAAGAGGGGGLGNOGAGRGGQGAAGAGGGGGY 152
QY 46 GLGSGGTSGRGGLGGGAGAAAAA-----AGAGGGGGLGSGGTSGRGGL 96
DB 153 GLGSGG-AGRGGLGGGAGAAAAAAGAGGGGGLGCGAGAGGGLGSGG-AGRGGL 210
QY 97 GGGGAGAAAAA-----AAAAAGAGGGGGLGSGGTSGRGGL 135
DB 211 GGGGAGAAAAAAGAGAGGGGGLGCGAGAGAGAGAGAGGGGGLGSGG-AGRG 266
QY 136 GGGGAGAAAAA-----AGAGGGGGLGSGGTSGRGGLGCGAGAAAA- 182
DB 267 GGGGAGAAAAAAGAGAGGGGGLGCGAGGGGGLGSGG-AGRGGLGCGAGAAAAAGAG 325
QY 183 -----AAAAAGAGGGGGLGSGGTSGRGGLGCGAGAAAA----- 220
DB 326 GGGGAGAGAGAGAGAAAAAAGAGGGGGLGSGG-AGRGGLGCGAGAGAAAAAGAGG 384
QY 221 -----AAAAAGAGAGGGGGLGSGGTSGRGGLGCGAGAAAA-- 259
DB 385 GGGGGLGSGAGRGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 443
QY 260 -----AAAAAGAGGGGGLGSGG 279
DB 444 GAGAGGGGGLGNOGAGRGGAGAGAGAGAGGGGGLGSGG 485

```

RESULT 2
 EBNL_EBV STANDARD; PRT; 641 AA.
 AC P03211;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE EBNA-1 NUCLEAR PROTEIN.
 GN BKRF1.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;

```

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Sachwell S.C., Seglin C.,
RA Tufnell P.S., Barrell B.G.;
RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=86259739; PubMed=3460083;
RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RT "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RT proteins: a probable transcriptional initiation site.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RA Petri L., Sample C., Kieff E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
RT latent infection nuclear proteins.";
RL Virology 176:563-574(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX MEDLINE=96006523; PubMed=7553871;
RA Bochkarev A., Barwell J.A., Pinetzer R.A., Furey W.F. Jr.,
RA Edwards A.M., Freppier L.;
RT "Crystal structure of the DNA-binding domain of the Epstein-Barr
RT virus origin-binding protein EBNA 1.";
RL Cell 83:39-46(1995).
CC - FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSACTIVATING FACTOR
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORP.
CC - SUBCELLULAR LOCATION: NUCLEAR. FREE IN THE NUCLEOPLASM, SOMEWHAT
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC WITH THE NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: V01555; CAA24816.1; -.
CC DR EMBL: M13941; AAA45889.1; -.
CC DR PIR: A03773; Q0BE31.
CC DR PIR: S33021; S33021.
CC DR PDB: 1VHI; 23-DEC-96.
CC DR TRANSFAC: T00211; -.
CC KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
CC 3D-structure.
CC FT DOMAIN 87 352 GLY/ALA-RICH.
CC SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

```

Query Match 41.1%; Score 610.5; DB 1; Length 641;
 Best Local Similarity 48.6%; Pred. No. 9.1e-24;
 Matches 143; Conservative 14; Mismatches 104; Indels 33; Gaps 10;

```

QY 2 RGSHHHHHNGSASGRGGLG-GGGAGAAAAAAGAGAGG-GGYGGLSGGTSGRGGLG 59
DB 83 KGTN-----GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
QY 60 GGGAGAAAAAAGAGAGAGG-GGYGGLSGGTSGRGGLGCGAGAAAAAAGAG--GA 115
DB 131 GAGGAGAGAGAGAGAGAGAGAGAGAGAG--GAGGAGAGAGAGAGAGAGAGAG 188
QY 116 GGGGGLGSGGTSGRGGLGCG- GAGAAAAAAGAGAGGGGGLGSGGTSGRGGLG 174
DB 189 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
QY 175 GAGAGAAAAAAGAGAGGGGGLGSGGTSG-----RGGLGCGAGAAAAAAGAGG 230
DB 245 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304

```



```

OY 231 GCGTGGGSGTSGRGLGCGAGCAAAAAAGAGCGGTGCGSQTSGIR 284
DB 305 AGGAGGAGAGGAGAGAGAGGCGG-----RGRGSGRGRGSGGGRGGR 354

RESULT 3
Y208_MYCTU 3
ID Y208_MYCTU STANDARD: PRT: 1901 AA.
AC 053553:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEMETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN RV3508 OR MTV023.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriales; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Baccocck K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Kirogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL022022; CNA17745.1;
CC DR HSSP: P19972; 1KVD.
CC DR Tuberculat; RV3508;
CC DR InterPro: IPR000084;
CC DR Pfam: PF00934; Pf.1.
CC KW Hypothetical protein; Repeat; Signal.
CC FT SIGNAL 1 30 POTENTIAL.
CC FT CHAIN 31 1901 HYPOTHEMETICAL PE-PGRS FAMILY PROTEIN
CC FT RV3508
CC FT SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;

Query Match 37.5%; Score 556.5; DB 1; Length 1901;
Best Local Similarity 44.9%; Pred. No. 6.6e-21;
Matches 140; Conservative 13; Mismatches 112; Indels 47; Gaps 14;

```

```

OY 170 GGLGCG--AGAAAAAAGAG--AGGCGTGLSGGT---SGRGLGCGAGCAAAAA 223
DB 593 GGAGCGGSSGAGGTNGSGAGGTGGGAGAGAGAGADNPTEIGAGGTGTGAAGAG 652

OY 224 AAAGAGGCGGCG-----LGSGTSGRGLGCG--GAGAAAAAAGAGAGAGAGAG 271
DB 653 GAGAGTGTGTGAGVSGVGNAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 712

OY 272 YGGLSGSGTSGT 283
DB 713 EGGAG--GNSGV 722

RESULT 4
ID SPD2_NEPCL STANDARD: PRT: 627 AA.
AC P46804;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 2 (DRAGLINE SILK FIBROIN 2) (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
NCBI_TaxID=6915;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92406876; PubMed=1527052;
RA Hman M.B., Lewis R.V.;
RT "Isolation of a clone encoding a second dragline silk fibroin.
RT Nephila clavipes dragline silk is a two-protein fiber."
RL J. Biol. Chem. 267:19320-19324(1992).
CC -1- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSERS UNIQUE
CC CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
CC PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSHEED
CC WITH ELASTIC AMORPHOUS SEGMENTS.
CC -1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M92913; AAA29381.1;
CC DR HSSP: P04002; 1WFA.
CC KW Silk; Repeat.
CC FT NON_TER 1 1
CC FT DOMAIN 1 530 15 APPROXIMATE TANDEM REPEATS.
CC FT REPEAT 1 36 1.
CC FT REPEAT 37 79 2.
CC FT REPEAT 80 121 3.
CC FT REPEAT 122 172 4.
CC FT REPEAT 173 213 5.
CC FT REPEAT 214 252 6.
CC FT REPEAT 253 283 7.
CC FT REPEAT 284 317 8.
CC FT REPEAT 318 359 9.
CC FT REPEAT 360 391 10.
CC FT REPEAT 392 428 11.
CC FT REPEAT 429 464 12.
CC FT REPEAT 465 488 13.
CC FT REPEAT 489 515 14.
CC FT REPEAT 516 530 15.
CC FT SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;

Query Match 37.3%; Score 554; DB 1; Length 627;
Best Local Similarity 44.1%; Pred. No. 4.6e-21;
Matches 166; Conservative 15; Mismatches 101; Indels 94; Gaps 17;

```

[illegible][illegible]

	RESULT	5
GRPL_PTHY	GRPL_PTHY	
ID	P09789	
AC	P09789	
NCBI_TaxID=4102;		
OX	NCBI_TaxID=4102;	
OC	Solanales; Solanaceae; Petunia.	
CC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae I;	
CC	Eukaryota; Viridiplantae).	
GN	GPR-1.	
OS	Petunia hybrida (Petunia).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
CC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;	
CC	Solanales; Solanaceae; Petunia.	
RP	SEQUENCE FROM N.A.	
RA	Condit C.M., Meagher R.B.:	
RT	"A gene encoding a novel glycine-rich structural protein of petunia."	
RL	Nature 323:178-181(1986).	
CC	-1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).	
CC	-1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).	
CC	-1- MISCELLANEOUS: THIS PROTEIN CONTAINS 67% GLYCINE RESIDUES.	
CC	-1- MISCELLANEOUS: 90% OF THE MATURE PROTEIN RESIDUES ARE CAPABLE OF FORMING A BETA-PLEATED SHEET COMPOSED OF 8 ANTI-PARALLEL STRANDS.	
CC	-1- MISCELLANEOUS: THE GLYCINE-RICH REGION IS COMPRISED OF TWO RELATED FAMILIES OF REPEATS, F1 AND F2, EACH REPEAT CONTAINING ABOUT 40 AA.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - - CC between the Swiss Institute of Bioinformatics and the EMBL outstation - - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/ or send an email to license@sib-ch).	
CC	EMBL; X04335; CAA27866.1; -.	
DR	PIR; A26099; A26099.	
DR	HSSP; P30129; ADPV.	
KW	Cell wall; Structural protein; Repeat; Signal.	
FT	SIGNAL	1 27
FT	CHAIN	28 384
FT		GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1.

CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; Z80225; CAB02341.1; ..
DR	HSSP; P04002; IATF.
DR	TubercuList; RV2634C; ..
DR	InterPro; IPRO000084; ..
DR	InterPro; IPRO01899; ..
CC	-----
OC	Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxId:1773;
RN	[1]
RX	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Holvansky T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	Nature 393:537-544(1998)."
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGNS
CC	SUBFAMILY.
CC	-----


```

OY 119 -----GTC-GLGSGGTGRCGLGCGAGAAAAAAGCAGGCGTGLGSGGTGRCG 171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4939 GYGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGS 4998
OY 172 LGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGS 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4999 GAGSACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGS 5058
OY 230 GCG-----GTC-GLGSGGTGRCGLGCGAGAAAAAAGCAGGCGTGLGSGGTGRCG 275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5059 GSGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGS 5118
OY 276 GSGGTGSG 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5119 GSGAGAG 5125

RESULT 8
GRP2_PHAUV STANDARD: PRT: 465 AA.
AC P10496;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 01-OCT-1994 (Rel. 30, Last annotation update)
GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1.8 PRECURSOR (GRP 1.8).
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fabiales; Fabaceae; Papilionoideae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TENDERGREEN;
RX MEDLINE=89091109; PubMed=3208742;
RA Keller B., Sauer N., Lamb C.J.;
RT "Glycine-rich cell wall proteins in bean: gene structure and
RT association of the protein with the vascular system.";
RL EMBL J. 7:3625-3633(1988).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -1- DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLATED
CC SHEET CONFIGURATION.
CC -1- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
CC ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
CC STRUCTURAL PROTEIN GRP 1.0.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X13596; CAN31932.1; -.
DR PIR: S01820; S01820.
DR HSP: P30129; 4DPV.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 465 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.8.
FT FT 33 465 GLY-RICH.
FT DOMAIN 205 359 8 X 22 AA TANDEN REPEATS.
FT FT 465 465 B5C4A9B983B45607 CRC64;
SQ SEQUENCE 465 AA; 36683 MW; B5C4A9B983B45607 CRC64;

Query Match 33.4%; Score 496.5; DB 1; Length 465;
Best Local Similarity 41.3%; Pred. No. 2.2e-16;
Matches 128; Conservative 8; Mismatches 139; Indels 35; Gaps 8;
OY 3 GSHHHHSSMASSRGGLGCGAGAAAAAAGCAGGCGTGLGSGGTGRCGLGCGG 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 GSHGIGYGGGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGS 190

```

```

OY 63 AGAAAAA-----AAAGCAGGCG--TGLGSGGTGRCGLGCGAGAAAAA----- 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 AGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 250
OY 110 -AAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 GAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 310
OY 163 SGTGSGRGGLGCGAGAAAAA-----AGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 AGGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 370
OY 217 AAAAAA-----AAAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 AGGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 430
OY 273 GGLGSGGTSG 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 GG-GAGGCGG 439

RESULT 9
YD25_MYCTU STANDARD: PRT: 603 AA.
AC Q10637;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV1325C PRECURSOR.
GN RV1325C OR MYCY130.10C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Mature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z73902; CA98089.1; -.
DR HSP: P19972; IKVD.
DR InterPro: IPR00084; -.
DR Pfam: PF00934; Pf. 1.
KW Hypothetical protein; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 603 HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT FT 114 603 GLY-RICH.
FT DOMAIN 603 603 RV1325C.
FT FT 49575 MW; 4F9BC82B07AE964 CRC64;
SQ SEQUENCE 603 AA; 49575 MW; 4F9BC82B07AE964 CRC64;

```


DR	PIR: S17732; KMMU.	
DR	HSP, P30129; ADPV.	
KW	Cell wall; Structural protein; Repeat; Signal.	
FT	SIGNAL	1 20
FT	CHAIN	21 338
FT		
FT	DOMAIN	21 338
SO	SEQUENCE	338 AA; 23891 MW; 046A6E8C1A4E89EB CRC64;

Query Match	33.0%;	Score 490;	DB 1;	Length 338;
Best Local Similarity	42.4%;	Pred. No. 3.8e-18;		
Matches 118;	Conservative 10;	Mismatches 140;	Indels 10;	Gaps 6;

QY	11	GSMASGRGGLCGDCAAGAAAAAAAAGAG---	-GCGTGG---	LDSCGTSGHGLGCGAGA	85
Db	58	GGLGGAGAGGGG	IGGAGAGCGAGGLGGGAGAGGLGCGHGGG	IGCGAGGAGAGGGLGCGHGG	117
QY	66	AAAAAAAAAGCAGCGCGTGTGCTG	SGTGTG	-GGCGAGAAAAAAAAGACGCGGTGTG	124
Db	118	IGGGAGGGGGGGLGCGGTGGGAGGAGGGG	IGCGHGGG	IGGAGAGGAGGGLGGGHHGGTG	177
QY	125	SGGTSGRGGTGGCGAGAAAAAAAAGCAGAGCGGTGTG	SGTGTG	SGTGTGCGAGAAAAAA	184
Db	178	--GCGAGGSGGGGLGCGGTGGAGGAGGAGGCGGAGGGG	GLCGHGGG	FGCGAGAGGLGAGAGCG	235
QY	185	AAAAGCAGCGGTGTG	SGTGTG	SGTGTG	244
Db	236	TGGGFGGAGGAGGAGGAGCGFGCGAGGAGCG	-GCGTGG-	-GAGGHHG	293
QY	245	RGGLGCGCAGAAAAAAAAGCAGCGGTGTG	SGTGTG		282
Db	294	GCGVGGFGGSGGGFGCGAGGAGGAGCGGTGGGGAG			331

RESULT	12	
Y747_MYCTU		
ID	Y747_MYCTU	STANDARD;
		PRT; 801 AA.

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE HYPOTHEMETICAL, PE-PGRS, FAMILY PROTEIN RV0747 PRECURSOR.
GN RV0747 OR MTY041.21..
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium
OX NCBI_taxid=1773;

RP SEQUENCE FROM N.A.
RC STRAIN=H37RV.
RX MEDLINE=96295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parish J., Garnier T., Churcher C., Harris D.
RA Gordon S.V., Eismelter K., Gass S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moulé S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajendram M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squires S., Squires R., Sulston J.E.
RA Taylor K., Whittread S., Barrell B.G.;
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the
RT complete genome sequence."; *Nature* 393:537-544(1998).
RL
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PERS
SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on lists on CC used by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/isb-sib.ch>) or send an email to license@isb-sib.ch).

CC	----
DR	EMBL; AL021958; CA17514.1; -.
DR	Tuberculist; RV0747; -.
DR	InterPro; IPR000084; -.
DR	Pfam; PF00934; PE; 1.
KW	Hypothetical protein; Repeat: Signal.
FT	SIGNAL
FT	1 30
FT	POTENTIAL.
FT	HYPOTHEICAL PE-PGRS FAMILY PROTEIN
FT	RV0747.
SQ	SEQUENCE 801 AA; 65407 MM; EAS4C9BFA5A00F41 CRC64;

Query Match	33.0%;	Score 490;	DB 1;	Length 801;
Best Local Similarity	38.6%;	Pred. No. 6.3e-18;		
Matches 133;	Conservative 14;	Mismatches 120;	Indels 78;	Gaps 14;

```

OY      11  GMSMSGRGGLGGGCGAGAGAAAAAAAAAAAAAAAAAGGA-----GDUbGtGtGdLsYUj:TSbGgUjLsUg- 0
Db      425  GGMFLPGNGNGNGCGCTTIGCGVNGAGCAGCAGAGGILFTGTGTGTGSGSGGPGATTGTGLGTGCGAGGAA 484
OY      63  -----AGAAAAAAAAAAAAAG-----AGGGGYGGLGSOGTSGRGGTGGCGGAGA 103
Db      485  LLFGSGGSGSGGAGAAVGGNGCGAGNGALLGAAAGAGCAGCAGACAVG--GNCAGAGNC-GL 541
OY      104  AAAAAAAAAAAG-----AGGGGYGGLGSG--GTSGRGGTGGGAGAAAAAAAAAAGC---AG 154
Db      542  FANGGAGPGGFGFSPAGAGGTGGAGGNGGLTFGAGGTGGAGGGSTTLGAGAGCAGGNGGLFG 601
OY      155  OGGYGGTGSOGT-----GRRGLGGO-----GAGAAAAAAAAAAGGA----- 191
Db      602  AGCTGTGACASHSTTAAGVSGAGCAGAGAGDAGLLSLGASGAGSGSGSSLTAAVYGGTGGAGC 661
OY      192  ---GGGGYGGGLGSOGTSGRGGTGGCGGAGA-----AAAAAAAAAAGAGGGGGYGGTGSOG 241
Db      662  LLFGSGGAGGSGSGFNSGNGCAGCAGGADAGLLVSGCAGCAGCAGATGAATGCGDGGAG--G 719
OY      242  TSGRGGTGGCGAGAAAAAAAAAAGAGAGCGGTGCG---LGSGGTSG 282
Db      720  KSGAGTGLGGDGGAGCATGTLGSAFHTIGGCGVGGSAVVLIGNGNGC 764

```

RESULT	13	
Y118_MYCTU		
ID	Y118_MYCTU	STANDARD;
		PRT; 498 AA

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV1818C.

OC Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devell K.R., Feltham T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rastandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulton J.E.

```

RT      deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence.";
RT      Nature 393:544(1998).
CC      -1. SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL)
CC      -1. SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PERS
CC      SUBFAMILY.
CC      -----

```


Job time: 732 sec

```
GN RV0278C OR MTW035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Besham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1 SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL021930; CAA17353.1; -.
CC DR Tuberculist: RV0278c; -.
CC DR InterPro: IPR000084; -.
CC DR Pfam: PF00934; Pfam: 1.
CC DR Hypothetical protein, Repeat, Signal.
CC KW SIGNAL
CC FT CHAIN 31 957
CC FT HYPOTHETICAL PE-PGRS FAMILY PROTEIN
CC FT RV0278C.
CC SQ SEQUENCE 957 AA; 81905 MW; 71EBABD417FBA47C CRC64;

Query Match 31.4%; Score 467; DB 1; Length 957;
Best Local Similarity 39.7%; Pred. No. 8.8e-17;
Matches 127; Conservative 12; Mismatches 129; Indels 52; Gaps 13;

QY 11 GSMASGRGGLGGGAGAAAAAAGCA-----GCGGYGGLGSGTSGRGGLGGGAGA 65
DB 570 GGMLEFNGNGNGGHGATNTATATGAGAGGILLFTGNGGTGTATF-GAGGIGAG-GA 627
QY 66 AAAAAAAGAGAGGCGGCGGCGTSGRGGLGGGCA--GAAAAAAGCAAG--AGCGGY 120
DB 628 GGVSVLLIGSGGTGNGNGNGSIGVAGIGGAGRGDAGLLFMAAGTGGHGAAGVPAGVGA 687
QY 121 GGLGSGGTSGRGGLGGGAGAAAAAAGAGAGCGGYGGLGSGQ---GTSGRGL--G 173
DB 688 GGNG--GLFANGAGAGAGGFNAAGNGNGGGLFTGTGTGAGTNGFAGGNGNGGGLFTGAG 745
QY 174 GCGAGAAAAAAGAGAGAGGCGGYGGLGSGTSG-----RGGLGGGAGAAAAA 222
DB 746 GTGGAGAGSGSGCTTTGGGGHGNAGLLSLGASGAGSGGASSLAGAGGTGNGALLFG 805
QY 223 AAAAGAG-----QGGYGLGSGQ---GTSGRGLGGGAGAAAAAAGCAAG 266
DB 806 FNGAGAGGHHGAGALLTSLQGGGAGAGNGGILLFSGAGAGAGAGGSGAN--ALGAGTGT 862
QY 267 AGCGGYGL-GSGGTSGTIR 285
DB 863 GGDGHHAGVFGNGGDCGCR 882
```

This Page Blank (uspto)

OY	110	AAAG-----GAGGGGTGGLGSQGTSSRGGTGGGAGAAAA-----	144
Dd	161	GGAAGGTGGTGGGAGAGGGTGGTGGSQG-AGRGTLGGGAGAAAAAAAAAAGAGGGTGGG	219
OY	145	-----AAAAAAGAAGCGCTGTGTSQGT-----	166
Dd	220	AGCGAGAAAAAAGAGAGGGTGTGTSQGAGRCGGAGAAAAAAVAGGGGTGGGAGGGY	279
OY	167	-----SGRGGTGGGAGAAA-----	182
Dd	280	GGTGGGAGRGGTGGGAGAAAAAAGAGGGTGGGAGAGCAGCAAAAAAGAGGGTGG	339
OY	183	-----AAAAAAGCAGCGTGTGTSQGTSSRGGTGGGAGCAAAAAA-----	225
Dd	340	LGNAGAGRCGGAAAAAAGGAGGGTGTGTSQG-AGRGTLGGGAGAAAAAAGAGGGY	398
OY	226	-----AGCAGGGGTGGTGGGTSSRGGTGGGAGAAAA-----	261
Dd	399	GGTGGGAGAGGGTGTGTSQG-SGRGGTGGGAGAAAAAAGAGGGTGGGAGGAGAAAA	457
OY	262	AAAGAGCGGTGGTGGSQG	279
Dd	458	AAAGGTGGGTGGTGGSQG	475

RESULT	2			
046171				
ID	046171	PRELIMINARY;	PRT;	544 AA.
AC	046171;			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
D7	01-MAR-2001 (TREMBLrel. 16, Last annotation update)			
DE	SPIDROIN 1 (FRAGMENT).			
OS	Nephila clavipes (orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila			
OX	NCBI_TaxID=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90384959; PubMed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 449-544 FROM N.A.			
RX	MEDLINE=98148687; PubMed=9487707;			
RA	Arcidiacono S., Mello C., Kaplan D., Cheley S., Bailey H.;			
RT	"Purification and characterization of recombinant spider silk			
RL	expressed in <i>Escherichia coli</i> .";			
RN	Appl. Microbiol. Biotechnol. 49:31-38(1998).			
RP	[3]			
RP	SEQUENCE FROM N.A.			
RA	Beckwilt R., Arcidiacono S., Stote R.;			
RL	Insect Biochem. Mol. Biol. 0:0-0(1998).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RA	Beckwilt R.;			
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U20329; AAC36957.1; .-			
SO	NON_TER			
FT	SEQUENCE 544 AA; 44107 MW; CC611B3551945615 CRC64;			

Query Match	67.98;	Score 1008;	DB 5;	Length 544;
Best Local Similarity	69.98;	Pred. No. 2e-62;		
Matches 230; Conservative	8;	Mismatches 17;	Indels 74;	Gaps 13;

```

QY      11  GSMASGRGGLGGGAGAAA-----AAAAAGAGGGGCTGGLSSG  51
      12  : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      64  GSGCGRGLGGGCGGAAAAGGCGGGLGGGAGAGCGAGAAAAGGAGGGGGLSSG  122
      65  : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      52  TSGRGLGGGCGAAAAAAAAAAGAGCGGCTGGLSSGCTSGRGLGGGAGAAAAAAA  111
      53  : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

RESULT	3		
ID	017434	PRELIMINARY;	PRT; 988 AA.
AC	017434;		
DT	01-JAN-1998 (TrEMBLrel. 05, Created)		
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last annotation update)		
DE	MINOR AMPULLATE SILK PROTEIN MISP1 (FRAGMENT).		
OS	Nephila clavipes (Ord spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.		
OX	NCBI_Taxid=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98200471; PubMed=9541398;		
RA	Colgin M.A., Lewis R.V.;		
RT	"Spider minor ampullate silk proteins contain new repetitive sequences		
RT	and highly conserved non-silk-like 'spacer regions'."		
RL	Protein Sci. 7:667-672(1998).		
DR	EMBL: AF027735; AAC14589.1; -.		
DR	InterPro: IPR000817; -.		
DR	PRINTS: PR00341; PRION.		
FT	NON_TER 1		
SQ	SEQUENCE 988 AA; 79082 MW; 461E03DF53F7085D CRC64;		

Query Match	45.58;	Score 675;	DB 5;	Length 988;
Best Local Similarity	45.48;	Pred. No. 2.4e-39;		
Matches 167; Conservative	6;	Mismatches 95;	Indels 100;	Gaps 12;

[illegible]

```

OY 272 YGGLSGSG 279
DB 598 YGGGGYGG 605

RESULT 4
ID 016988 PRELIMINARY: PRT: 410 AA.
AC 016988;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE FIBROIN-4 (FRAGMENT).
GN ADF-4.
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
OX NCBI_TaxID=45920;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9617878; PubMed=8600519;
RA Guerette P.A., Ginzinger D.G., Weber B.H., Gosline J.M.;
RT "Silk properties determined by gland-specific expression of a spider
RT fibroin gene family.";
RL Science 272:112-115(1996).
DR EMBL; U47836; AAC47011.1; -.
FT NON_TER 1
SQ SEQUENCE 410 AA; 34904 MW; E87A77BD59FF35C CRC64;

Query Match 43.1%; Score 640.5; DB 5; Length 410;
Best Local Similarity 56.9%; Pred. No. 2.6e-37;
Matches 169; Conservative 24; Mismatches 61; Indels 43; Gaps 16;

OY 14 ASGRCGLG--CGC-----AGAAAAAAAAAGAGAGCGGCGGCGTSGRGLGCG 62
DB 11 ASGCGTYPENCGSPGPAVPGCGPVSSAAAAAGSPGCGTGP-ENCGPSPGCGTGGC 69
OY 63 AGAAAAAAAAAGAGAGCGGCGGCGTSGRGLGCG-----GGA-----GAAAAAAA 111
DB 70 SSSSAAAAAAA--SGPGGYP-GSQGSPGCGSGYGPAGSGAGPGGPGASAAAAAAA 126
OY 112 AGGAGCGGCGGCGTSGRGLGCGGAGAAAAAGAGAGCGGCGGCGTSGRGLG 171
DB 127 AAASGPGGCGTGP-GSGCGPGAGYGPAGPGSSAAAAAAA--SGPGGYP-GSGPSPGCGV 182
OY 172 LCGCGAGAAAAAGAGAGCGGCGGCGTSGRGLGCGGAGAAAAAGAGAGAG 231
DB 183 YGPGGPGSSAAAAAAA--GSGCGTGP-ENCGPSPGCGTGPAGSGSSAAAAAAA--SGP 237
OY 232 GGGGCGGCGTSGRGLGCG-----GAGAAAAAGAGAGCGGCGGCGTSGTSG 282
DB 238 GGYGP-GSQGSPGCGGCGGCGGCGGAGAAAAAGAGAGCGGCGTSGTSG 292

RESULT 5
ID 076786 PRELIMINARY: PRT: 2639 AA.
AC 076786;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE FIBROIN.
OS Antheraea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Bombycoidea; Saturniidae; Saturniinae; Antheraea.
OX NCBI_TaxID=7119;
RN [1]
RP SEQUENCE FROM N.A.
RX Sezuan H., Tamura T., Yukuhiro K.;
RT "Characterization of the full length fibroin gene of a wild silkworm,
RT Antheraea pernyi.";
```

```

RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083334; AAC32606.1; -.
SQ SEQUENCE 2639 AA; 216056 MW; 2EE3310DEEB09B9A CRC64;

Query Match 41.6%; Score 617.5; DB 5; Length 2639;
Best Local Similarity 39.8%; Pred. No. 4.9e-35;
Matches 174; Conservative 10; Mismatches 90; Indels 163; Gaps 13;

OY 14 ASGRCGLG-----CGAGAAAAAGAGAGCGGCGGCGTSGRGLGCG 52
DB 1142 ASGAGCGGCGGCGGCGGCGGSDSAAAAAGAGAGCGGCGGCGGSDSAAAAAAA 1201
OY 53 -----SGRGLGCG-----CGAGAAAAAGAGAGCGGCGGCGTSGRGLGCG 90
DB 1202 AAAAGAGAGCGGCGGCGGCGGSDSAAAAAGAGAGCGGCGGCGGCGGSDS 1261
OY 91 -----SGRGLGCG-----CGAGAAAAAGAGAGCGGCGGCGTSGRGLGCG 121
DB 1262 AAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1321
OY 122 -----SGRGLGCG-----GAGAGCGGCGGCGGCGGCGGCGGCGGCGG 136
DB 1322 GSSAAAAARRAGHRAAGSAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGG 1381
OY 137 --CGAGAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 174
DB 1382 SAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1441
OY 175 -----CGAGAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 212
DB 1442 GCGSDSAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1501
OY 213 CGAGAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271
DB 1502 YGSDSAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1561
OY 272 YGGLSGCGTSGTGRPA 288
DB 1562 DGGYGS-GSSAAAAAAA 1577

RESULT 6
ID 002402 PRELIMINARY: PRT: 738 AA.
AC 002402;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE INSOLUBLE PROTEIN.
OS Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pteriolida;
OC Pteriolida; Pteriolida; Pteriolida.
OC Pteriolida; Pteriolida; Pteriolida.
OX NCBI_TaxID=50426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97320490; PubMed=9177341;
RA Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
RA Nakashima K., Takahashi T.;
RT "Structures of mollusc shell framework proteins.";
RL Nature 387:563-564(1997).
DR EMBL; D86074; BAA20466.1; -.
SQ SEQUENCE 738 AA; 61723 MW; FDF984139BF3BA59 CRC64;

Query Match 40.0%; Score 593.5; DB 5; Length 738;
Best Local Similarity 41.1%; Pred. No. 7.3e-34;
Matches 167; Conservative 10; Mismatches 82; Indels 147; Gaps 12;

OY 11 GSMAGRGGLGCGGAGAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 54
DB 113 GGMAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 172
```

```

QY 55 -----RGLGCGGAGAAAAAAGAGCGGCGTSGRGLGCG 99
DB 173 DLFDLFDLGAALALGAGAGCAAAAAAAGCGGAAAAAAGCGAGR 232
QY 100 --GAGAAAAAAGAGAG--OGCGGGLGSGTSGRGLGCG--AGAAAAA 150
DB 233 LGAAGAAAAAAGAGAGAGLGLGGLG--GLGGLGGLGGLGGSAAAAA 291
QY 151 GAGGCGGGLGSGTSGRGLGCGGAG-----AAAAAAGAGCGGCGGLG 203
DB 292 AAGGCGGLGCGTSG--GCGGRGRCGRGRRRAAAAAAAGCGGCGG-- 344
QY 204 TSGRGLGCGGAGAAAAA----- 226
DB 345 ---GGGCGGAAAAAASASASRONGIRDALDKILRSNGASAKASAY 400
QY 227 -----GAGCGGCGGLGSGTSGR-- 245
DB 401 STKSQIDDKLVKDLGLKSSASASASASAGCGGCGGCGGAGAL 460
QY 246 -----GGLGCGGAGAAAAAAGAGCGGCGGCGTSG 282
DB 461 AALAAAGAGGLGCGGCGGALAA--GAGGCGTGLGGLG 505

```

```

RESULT 7
ID 053559 PRELIMINARY; PRT: 1489 AA.
AC 053559;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE PGRS-FAMILY PROTEIN.
GN RV3514 OR MT0023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DB EMBL: AL022022; CAI17751.1; -
DR Tuberculist: RV3514; -
DR InterPro: IPR000084; -
DR InterPro: IPR002173; -
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223; -; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1489 AA; 115734 MW; 6855CBALIC3CBAP3A CRC64;

```

```

Query Match 37.4%; Score 556; DB 2; Length 1489;
Best Local Similarity 39.3%; Pred. No. 5e-31;
Matches 150; Conservative 17; Mismatches 103; Indels 112; Gaps 16;
QY 11 GSWASRGGLGCG--GAGAAAAAAGAGCGGCGGLGCG-----GTS 53
DB 292 GAGLGTGGTGGAGGAGGAGGALLGAGCGGGLGAGAGGCGTGGAGDGLGVGGTGG 351
QY 54 GAGGLGCG----GAGAAAAAAGAGCGGCGGGLGCGTSGRGLGCGA----- 101

```

```

DB 352 GAGCGVAGLGGAGGAGACQLFSASAGNAGYGCAGCGG--GDCGAGGADADDPGAT 409
QY 102 -----GAAAAAAGAG-----AGCGGCGGLGSGT-----SGRGLGCGGAGA 141
DB 410 GGTGFGAGGAGAGAGAGSSGAGGTGNSGGAGCGGAGGAGGAGADPRTIGGTGGGGTGG 469
QY 142 AAAAAAAGAGAGCGGCGG--LGSQGTSGRGLGCG-----GAGAA-- 182
DB 470 GAAAGAGAGAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 529
QY 183 AAAAAAGAG-----GCGGCGGLGSGG-----TSGRGLGCGGAGAGAAA 220
DB 530 GCGAGGAGAGSSGAGGTGNSGGAGGTGGGAGAGAGAGADNPRTIGGTGGGTGGAGA 589
QY 221 AAAAAAGAGCGGCGG--LGSQGTSGRGLGCG-----GAGAAAAA-----AAAA 264
DB 590 GAGGAGGAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 649
QY 265 GCGAGCGG---YGLGSGGTSG 282
DB 650 GCGAKAGGSSSAGGTNMSGAG 671

```

```

RESULT 8
ID 053552 PRELIMINARY; PRT: 1381 AA.
AC 053552;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE PGRS-FAMILY PROTEIN.
GN RV3507 OR MT0023.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DB EMBL: AL022022; CAI17744.1; -
DR HSSP: P00778; 20UL.
DR Tuberculist: RV3507; -
DR InterPro: IPR000084; -
DR InterPro: IPR002173; -
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223; -; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1381 AA; 110624 MW; CA09676BD07F6482 CRC64;

```

```

Query Match 36.7%; Score 544.5; DB 2; Length 1381;
Best Local Similarity 37.8%; Pred. No. 2.9e-30;
Matches 144; Conservative 17; Mismatches 115; Indels 105; Gaps 13;
QY 1 MRCSHHHHHSASRGGLGCGGAGAAAAAAGAGCGGCGGL--GSGCTSGRGL 58
DB 205 MGTGNGNGNGALLIGCGLG--GAGCMGCTGGGTGTGTCGNGNGALLIGAGCGVGGAGGI 262
QY 59 GCGGAGAAAAAAGAGAGAGCGGCGGGLGSGTSGRGLGCGGAGAAAAAAGAGAGAGAGAG 115
DB 263 GCGGTGAGGAGAGGTG--GNGAGGLFMNG--GDCGAGGCGGDDGAAADMAASAGGTGCK 318

```

```

QY 116 -----GCGGCGGLGSGQ-GTSGRGLGCGG-----AG 140
DB 319 GCGGCGDGTGTCAGGAGPVLFGHGGAGCGMGCGGTGCMGAGCGDGTIVIAAGTGGEGGTG 378
QY 141 AAAAAAAAAAGAGAGG--YGLGSGGCTSGRGLGCGGAGAAAAAAG-----GAGCG 194
DB 379 AAGAGGAAGARGALTSGLAGAGVAGGTGTGTGTGNGADAAAVYFGANGDPGFAGGK 438
QY 195 GYGLGSGQ-----GTSGRGLGCGQ-----GAGAAAAA 221
DB 439 GNGGIGGAANTGVAVAGCGGTGCGKGTGCGAGAGAGDAGSTGNPGKGGDGGIGAGAGGA 498
QY 222 AAAAA--AGGAGCGGYGGLGSGGCTSGRGLGCGQ-----GAGAAAAA 261
DB 499 AGTNGGAGHAAGTDCGCGCGGTGCGNGCGTGVNGADNTLNPDPFGAGEPGAGAGAGAG 558
QY 262 AAGGAGCGGYGGLGSGGCTSG 282
DB 559 AAGCGGTGTGTGNGCGNG 579

RESULT 9
ID 053557 PRELIMINARY; PRT: 1079 AA.
AC 053557;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PGRS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTV023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=9825987; PubMed=9634230;
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornbly T., Jagsels K., Krogan A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAI17749.1; -.
DR Tuberculist; RV3512; -.
DR InterPro: IPR002202; -.
DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
FT NON_TER 1 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CDBC74B97D CRC64;

Query Match 36.6%; Score 543.5; DB 2; Length 1079;
Best Local Similarity 40.7%; Pred. No. 2.7e-30;
Matches 144; Conservative 12; Mismatches 109; Indels 89; Gaps 14;
QY 14 ASGRGGLGCGA-GAAAAAAGAGAGGCGGYGGLGSGGCTSG-----GRG 56
DB 217 ANGIAGNGGDDGGAAGVATGATGAGDGHGTGAAGNGGTGAGGSGIDGVGGTG 276
QY 57 GLGGCGGAGAAAAAAGAGAG--GCGYGLGSGQ-----GTSGRGLGCGG- 100
DB 277 GTGCGNGGAGIGAGAGCGGSGNGGNGIGKGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 336
QY 101 -AGAAAAAAGAGAGAGCGGYGGLGSGGCTSGRGLGCGG-----AGAAAAA 148
DB 337 NGGAAGAAATGAGAGAGAGAGAG--GTGGRGSGGAGAGGAGGAGAGAGAGAGAGAGAG 394

```

```

QY 149 AAGAG-----GCGYGLGSGGCTSGRGLGCGGAGAAAAAAGAGAGCGYGL-- 199
DB 395 GAGAGGSGGPNTPSPGCGGCGGCGGCGAG-GAAGAGAGAGAGAGAGAGAGAGAGAGAG 453
QY 200 -----GSGCTSGRGLGCGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 246
DB 454 AGAASATNGSGGAGAGTGGGSGAGGTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 512
QY 247 GLGCGQ-----AGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 282
DB 513 GAGAGAGTGCNGNITGTAGTAGAAGNCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 566

RESULT 10
ID 016986 PRELIMINARY; PRT: 294 AA.
AC 016986;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE FIBROIN-2 (FRAGMENT).
GN ADF-2.
OS Araneus diadematus (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Araneidae; Araneus.
NCBI_TaxID=45920;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96178678; PubMed=8600519;
RA Guerette P.A., Ginzinger D.G., Weber B.H., Gosline J.M.;
RT Silk properties determined by gland-specific expression of a spider
RT fibroin gene family.";
RL Science 272:112-115(1996).
DR EMBL: U47854; AAC47009.1; -.
FT NON_TER 1 1
SQ SEQUENCE 294 AA; 24729 MW; BBCTE18AE2C9628 CRC64;

Query Match 36.5%; Score 541.5; DB 5; Length 294;
Best Local Similarity 52.3%; Pred. No. 1.2e-30;
Matches 136; Conservative 9; Mismatches 46; Indels 69; Gaps 10;
QY 11 GSMASGRGLGCGA-GAAAAAAGAGAGAGCGGYGGLGSGGCTSGR-----GLGGGAGAA 66
DB 1 GSGAGGAGAGCGGYGAGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 59
QY 67 AAAAAAAGC-AGCGYGLGSGGCTSGRGLGCGGAG-MAAAAAAAGAGAGCGGYGGLG 124
DB 60 AAAAAAGGCGGCGGCGGYGGLGSGGAGAGAGL-GYGAQGESAAAAAAGAGAGGCGGGLG 118
QY 125 SGTSGRGLGCGGAGAAAAAAGAGAGAGCGGYGGLGSGGCTSGRGLGCGGAGAAAAA 184
DB 119 -----AGAGAG-----YGAAGLGGCGGAGAGCGGSGAA 146
QY 185 AAGAGAGCGGYGGLGSGGCTSGRGLGCGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 244
DB 147 AAGAGCGGCGGYGGLGPGQ-----AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 177
QY 245 RGLGAGGAGAGAAAAAAGAA 264
DB 178 SLOYGGGGAOAAAAAASAA 197

RESULT 11
ID 053395 PRELIMINARY; PRT: 1538 AA.
AC 053395;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
RN PE_PGRS OR RV3345C OR MTV004.01C-MTV016.45C.

```

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RA Parkhill J.;
DR EMBL; AL021841; CAA17117.1; -;
DR HSSP; P00441; 1SOS.
DR Tuberculist; RV3345C; -;
DR InterPro; IPR000084; -;
DR InterPro; IPR002173; -;
DR Pfam; PF00934; PE; 1.1.
DR ProDom; PD001223; -; 1.
DR PROSITE; PS00583; PKB_KINASES_1; UNKNOWN_3;
SQ SEQUENCE 1538 AA; 129386 MW; 788F0B2095587592 CRC64;

Query Match 36.2%; Score 538; DB 2; Length 1538;
Best Local Similarity 38.9%; Pred. No. 8.8e-30;
Matches 126; Conservative 13; Mismatches 131; Indels 54; Gaps 9;

QY 11 GMSAGRGGLGCGAGAAAAAAGAGAGCGGCGTSGRGLGCGAGAAAAA 70
DB 1030 GNINGLGGFGAGGGGAVDAATTGCAAGNGAGGFASTGIGRGCAGCPGADFPAG 1089
QY 71 AAAAGCA---GQGGYGGIGLSQG-----TSGRGLGCGAGAAAAAAGG----- 114
DB 1090 VCGYVAGAGDGGGCGGCGGIGGEGRTGNGSGGCGGIGSLGNGGLGNGGVS 1149
QY 115 -----AGGGYGGIGLSQGTSGRGLGCGAGAAAAAAGAGCGGCGTSG 164
DB 1150 ETGFGAGGNGGTYGGP--GPEGNGGLGNGGAGNGGVSTTGDDGAGAGGNGDGCN 1207
QY 165 -----GTSGRGLGCGAGAAAAAAGG-----AGGCGYGGIGLSQGTSG 205
DB 1208 VGLGGDAGSGGAGNGGCTGTGAGAGAGAGAGNGSSKSTTTGNAAGSGAGNGTGIAN 1267
QY 206 GRGGLGCGAGAAAAA---AAAGCAGCGGTYGIGLSQGTSG--GRGGLGCGAGAAAAA 261
DB 1268 GAGGAGGAGGAGNAGVAGVAGGCGNGGNGGAGGDDTTGAGAGKGGNGSSGAASGSG 1327
QY 262 ---AAAGCAGCGGTYGIGLSQGTSG 282
DB 1328 VVNTVAGHGNGGNGGNGGAGSAG 1351

RESULT 12
ID 053844 PRELIMINARY; PRT; 749 AA.
AC 053844;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN RV0833 OR MTV043.25.

OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RA Parkhill J.;
DR EMBL; AL022004; CAA17639.1; -;
DR Tuberculist; RV0833; -;
SQ SEQUENCE 749 AA; 57448 MW; C3C91307F9A8A2EC CRC64;

Query Match 36.0%; Score 535; DB 2; Length 749;
Best Local Similarity 39.9%; Pred. No. 7.7e-30;
Matches 133; Conservative 9; Mismatches 123; Indels 68; Gaps 11;

QY 16 GRGGLGCGAGAAAAAAGAGAGCGGTYGIGLSQ-----GTSGR 55
DB 3 GNGAGGSGAGAPATGAGGAPAGLIGVGACAGAGDSNAGYVAGAGAGAAALFGAGGA 62
QY 56 GGLGCGAGAAAAAAGCA-----GQGGYGGIGLSQGTSGRGLGCGAGAAAAA 108
DB 63 GAGGSG--GSGAAGAGAGAGAGGLFASGSGSGFCGFASTGCGAGGTGAGAGLFPASGV 121
QY 109 AAAAGCAGCGTYGIGLSQG-----TSGRGLGCGAGAAAAAAGAGCGGCGTSG 161
DB 122 GGTGGAGSGGTGGYGTGAGAGLFPASGAGAGSGGTGGAGGTGAGAGLFPAGAGAGL 181
QY 162 GSQG--TSGRGLGCG-----GAGAAAAAAGAGCA--GQGGYGG--GSQGTSGR 207
DB 182 GCGNHTGCHGAGGSAGLALDGGAGAGAGAAATTGTGAGAGAGKAGLFGSGGAGGS 241
QY 208 GGLGCG-----GAGAAAAAAGAGAGCGGTYGIGLSQGTSGRGLGCGAGCA---- 255
DB 242 GGAAGTFCDTGNSGAGAGAGKAGLLFSGCAGGSGGAGGFANGSTGGAGAGCGAGLIG 301
QY 256 -----AAAAAAGAGAGCGGTYGIGLSQGTSGI 283
DB 302 NGGNGSGGTGATGAGNGAGAGAG--GGAAGL 332

RESULT 13
ID 093119 PRELIMINARY; PRT; 421 AA.
AC 093119;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ANTHEAERA PERNYI FIBROIN (FRAGMENT).
OS Antheraea pernyi (Chinese oak silk moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Bombycoidea; Saturniidae; Saturniidae; Antheraea.
OX NCBI_TaxID=7119;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POSTERIOR SILKGANDS;
RA Yukuhiro K., Kanda T., Tamura T.;
RT "Preferential codon usage and two types of repetitive motifs in the
RT fibroin gene of the Chinese oak silkworm, Antheraea pernyi1.";
RL Insect Mol. Biol. 0:0-0(1996).

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:06 ; Search time 86.32 Seconds
(without alignments)
478.278 Million cell updates/sec

Title: US-09-490-291-4
Perfect score: 3493
Sequence: 1 MRGSHHHHHMSMASGRGL.....GCGPGQOTSGIRPRAKLN 681

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	651	20	AAV40097
2	2135	61.1	718	19	AAW53346
3	2135	61.1	718	12	AAV59070
4	2126	60.9	718	21	AAV4308
5	2076	59.4	606	16	AAW9053
6	2076	59.4	606	20	AAV40100
7	2052.5	58.8	646	18	AAW27178
8	1972.5	56.5	606	16	AAW9055
9	1972.5	56.5	606	20	AAV40102
10	1971.5	56.4	604	16	AAW9057
11	1963.5	56.2	606	20	AAV40101

Result No.	Score	Query Match	Length	DB ID	Description
12	1405.5	40.2	831	16	AAW80168
13	1395	39.9	615	20	AAV40099
14	1331.5	38.1	714	16	AAW9059
15	1331.5	38.1	714	20	AAV40103
16	1233.5	35.3	1177	9	AAW80940
17	1233.5	35.3	1177	11	AAW80307
18	1233.5	35.3	1177	17	AAW85105
19	1233.5	35.3	1177	18	AAW26342
20	1233.5	35.3	1177	21	AAW78277
21	1233.5	35.3	1177	22	AAW72725
22	1233.5	35.3	1177	22	AAW63995
23	1233.5	35.3	1178	14	AAW41007
24	1233.5	35.3	1178	19	AAW53518
25	1176.5	33.7	1023	19	AAW53524
26	1176.5	33.7	1059	14	AAW41013
27	1176.5	33.7	1059	18	AAW26348
28	1176.5	33.7	1101	21	AAW78283
29	1172.5	33.6	1059	9	AAW82962
30	1166	33.4	980	16	AAW81318
31	1157	33.1	728	19	AAW56163
32	1132	32.4	595	12	AAW14309
33	1132	32.4	595	19	AAW53347
34	1132	32.4	595	21	AAW59071
35	1131	32.4	1332	17	AAW95109
36	1131	32.4	1332	22	AAW72729
37	1131	32.4	1332	22	AAW63999
38	1129	32.3	1038	17	AAW95107
39	1129	32.3	1038	22	AAW72727
40	1129	32.3	1038	22	AAW63997
41	1127	32.3	979	21	AAW78286
42	1127	32.3	1050	18	AAW26350
43	1114	31.9	1066	11	AAW05312
44	1105.5	31.6	1018	16	AAW01496
45	1104.5	31.6	531	20	AAV40098

ALIGNMENTS

RESULT 1
ID AAV40097 standard; protein; 651 AA.
XX AAV40097;
XX AC AAV40097;
XX DT 19-NOV-1999 (first entry)
XX DE Spider silk protein spiderone major 1.
XX KW Spider silk protein; spiderone major 1; cosmetic; make-up;
XX KW dermatological compositions; hair care; skin care; sunscreen;
XX KW hormone; moisturizer; skin disorder; skin disorder.
XX OS Nephila clavipes.
XX PN FR2774588-A1.
XX PD 13-AUG-1999.
XX PF 11-FEB-1998; 98FR-0001614.
XX PR 11-FEB-1998; 98FR-0001614.
XX (OREA) L'OREAL SA.
XX PI Philippe M, Garson JC, Arraudeau JP;
XX WPI; 1999-510729/43.
XX DR Cosmetic or dermatological composition containing spider silk protein,
XX PT for hair or skin care, in make-up or sunscreens -
XX PS Claim 3; Fig 1; 32pp; French.

XX The present sequence represents the natural spider silk protein
CC spidroine major 1. The protein improves the moisturizing/softening
CC action of the compositions. The protein, and its fragments are used
CC in cosmetic or dermatological compositions. These compositions have
CC use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.

XX Sequence 651 AA:

Query Match 61.1%; Score 2135; DB 20; Length 651;
Best Local Similarity 68.3%; Pred. No. 1.4e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
23 OGAGAAAAAAGAGAGGCGGCGGCGT--SGRGLGCGAG-AAAAAAAAAGAGAG 79
||||| ||||||| ||||||| || ||||||| ||||||| ||||||| |||||||
Db 1 ggaq-----aaaaaagagaggggylgggagggggylgggaggaagaaaaaagag 56
80 GGTGGLSGGTSGRGLGCGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGCGAG 139
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 57 ggygglgsgq---agrgggag---aaaaaagaggggylgggag--agrgglgggga 107
140 GAAAAAAGAGAGAGGCGGCGGCGTSGPGGCGGCGGCGGCGGCGGCGGCGG 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 g---aaaaaagaggggylgggag-----agrgggg-----aaaaa 142
200 AGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 143 aggaaggggylgggag--agrgglgggagaaaaaagaggggylgggaggggylg 201
251 SGTGSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 289
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 202 sgg--agrgglgggagaaaaaagagggylgggagggagaaaaaagagggylg 260
290 QGTSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 348
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 261 gq---agrgggag---aaaaaagaggggylgggag--aggggylgggagggyl 311
349 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 390
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 312 ggggagaaaaagagggylgggagggagaaaaaagagggylgggag--agrgglgggag 370
391 AGAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 450
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 371 ag---avaaaagaggggylgggag---agrgggag---aaaaaagagggylgggag 419
451 LGSQGTSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 497
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 420 lgnng--agrgglgggagag---aaaaaagaggggylgggagggagrggagaaagaggg 475
498 GGTGP--GQGTSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 476 ggygglgggagag---gggagaaaaaagagggylgggaggggylgggag--agrggl 531
551 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 590
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 532 ggggagaaaaaagagggylgggagggagaaaaaagagggylgggag---agrggg 587
591 CGAGAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 650
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 588 ggaq-----aaaaaagaggggylgggagv--grgglgggag-----aaaaaggggggyl 636
651 GGLGS 655
|||||
Db 637 ggvgs 641
```

RESULT 2

AAW53346
ID AAW53346 standard; Protein; 718 AA.

AC AAW53346;

DT 06-JUL-1998 (first entry)

DE Nephila clavipes spider silk protein.

XX Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;

KM cocoon; tensile strength; elasticity.

XX Nephila clavipes.

PN US5728810-A.

PD 17-MAR-1998.

PE 19-APR-1995; 95US-0425069.

PR 15-APR-1991; 91US-0684819.

PR 20-APR-1990; 90US-0511792.

PR 04-OCT-1994; 94US-0317844.

PR 19-APR-1995; 95US-0425069.

PA (UTWT-) UNIV WYOMING.

PI Human MB, Lewis RV, Xu M;

PS WPI: 1998-270437/24.

DR N-PSDB: AAV23249.

XX Recombinant spider silk proteins - useful for making fibres

XX Claim 1; Column 29-34; 68pp; English.

XX The present sequence represents a spider silk protein from the present

CC invention. Spider silk proteins, and peptide fragments within the

CC proteins, can be produced and purified independently and can then be

CC mixed and made into fibres that have higher tensile strengths and

CC elasticity than naturally occurring fibres. The fibres can be used in

CC mixed composites. The invention allows the two naturally occurring

CC Nephila clavipes silk proteins to be produced independently so that

CC they can later be combined to form silk fibres of high tensile strength

CC and elasticity.

CC Sequence 718 AA:

SQ

Query Match 61.1%; Score 2135; DB 19; Length 718;
Best Local Similarity 68.3%; Pred. No. 1.5e-148;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```
23 OGAGAAAAAAGAGAGGCGGCGGCGT--SGRGLGCGAG-AAAAAAAAAGAGAG 79
||||| ||||||| ||||||| || ||||||| ||||||| ||||||| |||||||
Db 1 ggaq-----aaaaaagaggggylgggagggggylgggaggaagaaaaaagag 56
80 GGTGGLSGGTSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 139
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 57 ggygglgggag---agrgggag---aaaaaagaggggylgggag--agrgglgggga 107
140 GAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 199
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 108 g---aaaaaagaggggylgggag-----agrgggg-----aaaaa 142
200 AGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 143 aggaaggggylgggag--agrgglgggagaaaaaagaggggylgggaggggylg 201
251 SGTGSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 289
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 202 sgg--agrgglgggagaaaaaagagggylgggagggagaaaaaagagggylg 260
```


KW protein superfibre: major ampullate silk; orb web spider.
 OS Nephilia clavipes.
 XX EP452925-A.
 XX PD 23-OCT-1991.
 XX PF 18-APR-1991; 91EP-0106217.
 XX PR 20-APR-1990; 90US-0511792.
 XX (UYWY-) UNIV OF WYOMING.
 XX Lewis RV, Xu M, Hinman M;
 XX WPI: 1991-312199/43.
 XX N-PSDB; AAO14183.
 DR DNA encoding spider silk protein-1 and 2 and variants - isolated
 PT from Nephilia clavipes, for prodn. of spider silk protein and
 PT fibres having desired characteristics
 XX
 PS Claim 15; Page 23; 48pp; English.
 CC The spider silk protein contains a basic 34 amino acid repeat. The
 CC repeat itself contains 3 regions. The first comprises 0-9 amino
 CC acids with a sequence AGR(GGX)2. This region is not highly
 CC conserved. The second region has a sequence GAG(A)x which is highly
 CC conserved and is 8-10 amino acids long. The third segment is (GGX)5
 CC and is 15 amino acids long and is very highly conserved. In most
 CC cases x is A,Q, Y or L. Removal of the poly-(Ala) segments results
 CC in a silk having lower elasticity.
 XX
 SQ Sequence 718 AA;
 Query Match 60.9%; Score 2126; DB 12; Length 718;
 Best Local Similarity 68.4%; Pred. No. 6,7e-148;
 Matches 492; Conservative 16; Mismatches 47; Indels 164; Gaps 34;
 QY 23 OGAGAAAAAAGAGAGGCGGCGGCGT--SGRGCGGCGGAG-AAAAAAGAGAGG 79
 DB 1 qgaq----aaaaaag 56
 QY 80 GGYGGLGSGQTSGRGGLGCGAGAAAAAAGAGAGGCGGCGGCGTSGRGGLGCGAG 139
 DB 57 ggygglgsgq----agrggagag----aaaaaagagagagagagagagagagag 107
 QY 140 GAAAAAAGAGAGGCGGCGGCGT-----SGPGGYGP-GQOTSGRGGL 184
 DB 108 g----aaaaaag 164
 QY 185 GCGGAGAAAAAAGAGAGGCGGCGGCGTSGRGGLGCGGAGAAAAAAG 235
 DB 165 ggggag 220
 QY 236 AAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 295
 DB 221 aagagagq-----gglgggagagagagagagagagagagagagagagag 262
 QY 296 GGLGCGGAGAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 354
 DB 263 agrggagag-----aaaaaagagagagagagagagagagagagagagag 317
 QY 355 AAAA-----AAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 396
 DB 318 aaaaag 373
 QY 397 AAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 456
 DB 374 vaaaaaag 424

QY 457 SGRGGLGCGAGAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 502
 DB 425 agrgglgagagag-----aaaaaagagagagagagagagagagagagagag 481
 QY 503 GQOTSGRGGLGCGAGAAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 556
 DB 482 gsggagag-----gggagagagagagagagagagagagagagagagagagag 537
 QY 557 AAAA-----AAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 596
 DB 538 aaaaag 591
 QY 597 AAAAAAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 655
 DB 592 -aaaaaag 641
 RESULT 5
 ID AAR99053 standard; Protein; 606 AA.
 XX AAR99053;
 AC AAR99053;
 XX
 DT 17-JAN-1997 (first entry)
 XX
 DE Spider dragline variant, DP-1A.9 polymer.
 XX
 KW Spider dragline protein; variant; monomer; polymer;
 KW fibre forming region; Spidroin 1; Nephilia clavipes; DPL; mimic;
 KW DP-1A analogue; fibre; high tensile strength; elasticity; clothing;
 KW rope; surgical suture; implant; reinforcement; film; coating.
 XX
 OS Synthetic.
 XX
 PN WO9429450-A2.
 XX
 PD 22-DEC-1994.
 XX
 PF 15-JUN-1994; 94WO-US06689.
 XX
 PR 15-JUN-1993; 93US-0077600.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX
 PI Farnestock SR;
 XX
 DR WPI: 1995-036479/05.
 XX
 PT New synthetic variants of spider dragline protein - for making
 PT fibres useful as clothing, surgical silk, plastic reinforcement
 PT etc., also related DNA, vectors and transformed cells
 PS
 XX Disclosure; Page 85-87; 168pp; English.
 CC
 CC This sequence represents a synthetic spider dragline variant polymer,
 CC DP-1A.9. The sequence of the DP-1A.9 monomer is given in AAR99052.
 CC The polypeptide monomer is a variant based on a consensus sequence
 CC derived from the fibre forming regions of spider dragline protein,
 CC esp. the natural protein 1 (Spidroin 1) from Nephilia clavipes.
 CC DNA sequence encoding the monomer may be used in the recombinant
 CC production of the variant protein in a recombinant host, e.g. E. coli
 CC or Bacillus subtilis. Synthetic analogues of DPL were designed to mimic
 CC the repeating consensus sequence of the natural protein and the pattern
 CC of variation among individual repeats. DP-1A analogues are composed
 CC of a tandemly repeated 101 amino acid monomer which comprises four
 CC repeats which differ from the consensus sequence given in AAR90201,
 CC according to the pattern (1)-(5) given below. This 101 amino acid
 CC monomer is repeated 1-16 times in a series of analogue proteins. The
 CC individual repeats differ from the consensus according to the pattern:
 CC (1) the poly-alanine sequence varies in length from 0-7
 CC residues; (2) when the entire poly-alanine sequence is deleted,
 CC so also is the surrounding sequence encompassing AGRGGLGCGAGAGG;
 CC (3) aside from the poly-alanine sequence, deletions usually

CC encompass integral multiples of three consecutive residues:
CC (4) deletion of GYG is generally accompanied by deletion of GRC
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.
XX
XX
SQ Sequence 606 AA:

Query Match 59.4%; Score 2076; DB 16; Length 606;
Best Local Similarity 69.7%; Pred. No. 2.5e-144;
Matches 477; Conservative 11; Mismatches 74; Indels 122; Gaps 30;

OY 19 GLGGGAGAAAAAAGAGGCGGCGTSGRCGLGGGAGAAAAAAGAGAG 78
DB 3 grggrgag-----aaaaagagaggggylgsgg-agrgylgsggag-----aaaaagagag 54
OY 79 OGCGYGLGSGCTSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLG 136
DB 55 q-----gglgsggagagagagagagaggggylgsggaggggylgsg 100
OY 137 GGA-----GAAAAAAGAGGCGGCGTSGRCGLG 185
DB 101 ggggrgagagagagagaggggylgsggagrglgggagagagagagagggylg 160
OY 186 GCGAGAAAAAAGAGGCGGCGTSGRCGLGGGAG-----GAAAAAAGAG 240
DB 161 aggagagagagagagaggggylgsggaggggylgsggagagagagagag 220
OY 241 AGCGGCGGCGTSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLG 300
DB 221 aggggylgsgg-agrgylgsggag-----aaaaagagag-----gglgsg 262
OY 301 OGAGAAAAAAGAGGCGGCGTSGRCGLGGGAGAGAGAGAGAGAGAGAGAG 359
DB 263 gggagagagagagagaggggylgsgg-aggggylgsggag-----gggag-----a 314
OY 360 AAAAAAGAGGCGGCGTSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLG 419
DB 315 aaaaagaggggylgsgg-agrgylgsggag-----aaaaagagag----- 358
OY 420 GRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLGGGAG-----GAAAAA 474
DB 359 --gglgsggagagagagagagaggggylgsggaggggylgsggagrgggagag 416
OY 475 AAAAAAGGCGGCGTSGRCGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAG 526
DB 417 aaaaagaggggylgsggagrglgggagagagagagagagagagagagagag 476
OY 527 AAGGAGGCGGCGTSGRCGLGGGAG-----GAAAAAAGAGGCGGCGTSGRCGLG 581
DB 477 aagagaggggylgsggaggggylgsggagrglgggagagagagaggggylgsgg 536
OY 582 TSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLGGGAGAAAAAAGAG 641
DB 537 -agrgylgsggag-----aaaaagagag-----gglgsggagagagagagag 578
OY 642 AGGAGGCGGCGTSGRCGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
DB 579 aggaggggylgsgg-aggggylgsgg 601

RESULT 6
AA140100
ID AA140100 standard; protein; 606 AA.
XX
AC AA140100;
XX
DT 19-NOV-1999 (first entry)

XX
DE Polymer of an analogue of spider silk protein spidroine major 1.
XX
KW Spider silk protein; spidroine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
KW hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephila clavipes.
XX
FH Key Location/Qualifiers
FT Peptide 1..101
FT /note="monomer unit"
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PF 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudeau JP;
XX
PT WPI; 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PT for hair or skin care, in make-up or sunscreens -
XX
PS Claim 8; Fig 4B; 32pp; French.
XX
CC The present sequence represents a polymer of an analogue of the spider
CC silk protein spidroine major 1. The protein improves the moisturizing/
CC softening action of the compositions. The protein, and its fragments
CC are used in cosmetic or dermatological compositions. These compositions
CC have use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.
XX
SQ Sequence 606 AA:

Query Match 59.4%; Score 2076; DB 20; Length 606;
Best Local Similarity 69.7%; Pred. No. 2.5e-144;
Matches 477; Conservative 11; Mismatches 74; Indels 122; Gaps 30;

OY 19 GLGGGAGAAAAAAGAGGCGGCGTSGRCGLGGGAGAAAAAAGAGAG 78
DB 3 grggrgag-----aaaaagagaggggylgsgg-agrgylgsggag-----aaaaagagag 54
OY 79 OGCGYGLGSGCTSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLG 136
DB 55 q-----gglgsggagagagagagagaggggylgsggaggggylgsg 100
OY 137 GGA-----GAAAAAAGAGGCGGCGTSGRCGLG 185
DB 101 ggggrgagagagagagaggggylgsggagrglgggagagagagagagggylg 160
OY 186 GCGAGAAAAAAGAGGCGGCGTSGRCGLGGGAG-----GAAAAAAGAG 240
DB 161 aggagagagagagagaggggylgsggaggggylgsggagagagagagagag 220
OY 241 AGCGGCGGCGTSGRCGLGGGAGAAAAAAGAGGCGGCGTSGRCGLG 300
DB 221 aggggylgsgg-agrgylgsggag-----aaaaagagag-----gglgsg 262
OY 301 OGAGAAAAAAGAGGCGGCGTSGRCGLGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359

DR WPI: 1999-510729/43.
 XX Cosmetic or dermatological composition containing spider silk protein,
 PT for hair or skin care, in make-up or sunscreens -
 XX
 XX Claim 8; Fig 6B; 32pp; French.
 XX
 CC The present sequence represents a polymer of an analogue of the spider
 CC silk protein spidroine major 1. The protein improves the moisturizing/
 CC softening action of the compositions. The protein, and its fragments
 CC are used in cosmetic or dermatological compositions. These compositions
 CC have use as hair or skin care products; and make-up or sunscreens.
 CC As the protein is a good, persistent film-formers on the skin
 CC of low surface density, it can be used for delivery of active
 CC agents that are generally difficult to administer, e.g. vitamins,
 CC hormones, moisturizers or agents for treating disorders of the
 CC skin and hair.
 XX
 XX Sequence 606 AA:

Query Match 56.5%; Score 1972.5; DB 20; Length 606;
 Best Local Similarity 64.2%; Pred. No. 8,8e-137;

Matches 462; Conservative 8; Mismatches 35; Indels 215; Gaps 29;

```

OY 38 GAGGCGGCGGCGTSGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLG 97
   ||||||||| :||||||| ||||||||| |||
Db 1 gagggggylggag--agrglgggag--aaaaaaggag-----gllg 42
OY 98 GCGAGAAAAAGAGAGCGGCGTSGRCGLGCGGAGAAAAAGAGCGG 157
   ||||| :||||||| ||||||||| |||||||||
Db 43 sggagggagaaaaaggagggylggag--agrgggag--aaaaaaggagggg 94
OY 158 YGGLSGCGTSGGCGYGP-GQGTSGRCGLGCGAGAAAA-----A 195
   ||||||| :||| | | :||||||| |||||
Db 95 ygglggag--aggggylggagagrglggaggaagaaaggagggylggaggaagaga 153
OY 196 AAAAGAGCGGCGTSGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLG 254
   ||||||||| :||||||| ||||||||| |||||||||
Db 154 aaaaaggagggylggag--agrgggag--aaaaaaggagggylggaggg 205
OY 255 -----SGRCGLGCGGAGAAAA-----AAAAAGAGCGG 282
   :||||||| ||||||||| ||||||||| |||||||||
Db 206 gggggylggagagrglggaggaagaaaggagggylggaggaagaaaggaggg 265
OY 283 GYGGLSGCGTSGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLGCGG 341
   ||||||| :||||| ||||||||| ||||||||| :||| | |
Db 266 gggglggag--agrgggag--aaaaaaggagggylggag--agggggylggag 316
OY 342 TSGRCGLGCGAGAAAA-----AAAAAGAGCGGCGTSGRCGLG 380
   :||||||| ||||||||| ||||||||| |||||||||
Db 317 ggggryglggaggaagaaaggagggylggaggaagaaaggagggylggaggg 375
OY 381 SGRGGLGCGGAGAAAAAGAGCGGCGTSGRCGLGCGG-----SGRCGLGCGG 428
   ||||||| :||||||| ||||||||| |||||||||
Db 376 ---agrgggag--aaaaaaggagggylggaggaagggylggaggaaggg 428
OY 429 AGAAAAAGAGAGCGGCGTSGRCGLGCGGAGAAAAAGAGCGGCGG 488
   || :||||||| ||||||||| ||||||||| |||||||||
Db 429 ag--aaaaaaggag-----gglggaggaagaaaggagggylggag 471
OY 489 LGSQGTSGRCGYPGQGTSGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLG 544
   ||||| :||||| ||||||||| ||||||||| |||||||||
Db 472 lggag-----agrgggag--aaaaaaggagggylggaggaaggg 511
OY 545 -----SGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLGCGGAGAA 596
   :||||||| ||||||||| ||||||||| ||||| |||||
Db 512 ygglggagggylggag--aaaaaaggag-----gglggaggaaggg 554
OY 597 AAAAAAGAGAGCGGCGTSGRCGLGCGAGAAAAAGAGCGGCGTSGRCGLG 656
   ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 555 agaaaaaggagggylggag-----agrgggag--aaaaaaggagggylggag 606

```

RESULT 10

AAR99057
 ID AAR99057 standard; Protein; 604 AA.

AC AAR99057;

DT 17-JAN-1997 (first entry)

DE Spider dragline variant, DP-1B.16 polymer.

XX Spider: dragline protein; variant; monomer; polymer;

XX fibre forming region; Spidroin 1; Nephlia clavipes; DP1; minic;

XX DP-1A analogue; fibre; high tensile strength; elasticity; clothing;

XX rope; surgical suture; implant; reinforcement; film; coating.

OS Synthetic.

XX W09429450-A2.

XX 22-DEC-1994.

XX 15-JUN-1994; 94MO-US06689.

XX 15-JUN-1993; 93US-0077600.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX Fahnestock SR.

XX WPI: 1995-036479/05.

XX New synthetic variants of spider dragline protein - for making

XX fibres useful as clothing, surgical silk, plastic reinforcement

XX etc., also related DNA, vectors and transformed cells

XX Disclosure: Page 106-108; 168pp; English.

XX This sequence represents a synthetic spider dragline variant polymer,

XX DP-1B.16. The sequence of the DP-1B.16 monomer is given in AAR99056.

XX The polypeptide monomer is a variant based on a consensus sequence

XX derived from the fibre forming regions of spider dragline protein,

XX esp. the natural protein 1 (Spidroin 1) from Nephlia clavipes. The

XX DNA sequence encoding the monomer may be used in the recombinant

XX production of the variant protein in a recombinant host, e.g. E. coli

XX or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic

XX the repeating consensus sequence of the natural protein and the pattern

XX of variation among individual repeats. This monomer exhibits all of the

XX regularities of (1)-(5) below. In addition, it exhibits a regularity of

XX the natural sequence which is not shared by DP-1A, namely that a repeat

XX in which both GYG and GNG are deleted is generally preceded by a repeat

XX lacking the entire poly-alanine repeat, with one intervening repeat.

XX The sequence of DP-1B matches the natural sequence more closely over

XX a more extended segment than does DP-1A. The individual repeats differ

XX from the consensus sequence given in AAW06201 according to the pattern:

XX (1) the poly-alanine sequence varies in length from 0-7

XX residues; (2) when the entire poly-alanine sequence is deleted,

XX so also is the surrounding sequence encompassing AGRGGLGCGAGANGG;

XX (3) aside from the poly-alanine sequence, deletions usually

XX encompass integral multiples of three consecutive residues;

XX (4) deletion of GYG is generally accompanied by deletion of GNG

XX in the same sequence; and

XX (5) a repeat in which the entire poly-alanine sequence is

XX deleted is generally preceded by a repeat containing six alanine

XX residues.

XX The proteins may be used to produce fibres of high tensile strength and

XX elasticity, suitable for clothing, rope, surgical sutures, biomaterials

XX for implants, plastic reinforcements, films, coatings, etc.

XX Sequence 604 AA:

XX

Query Match 56.4%; Score 1971.5; DB 16; Length 604;

OY		LGSOSTSRPGVGPQGTSGRGGTGLGCGAGAAAAAAGAAGCAGCGTGCTLSOGT----	544
Dd	474	Iysgg-----agrgggaq----aaaaaaaggaggyggIysggaqqg	513
OY	545	-SGRGCLGCQGCAAAAAGACAGCGYGTLCSOGTSCGCLGCGAGAA	596
Dd	514	yggylgsyggaqr.ygglgsyggaq---aaaaaaaaaqgr-----ggl.gsyggaqqg	556
OY	557	AAAAAAAAAGAGAGCGGYGCLGSQGTSGRGCLGCGAGAAAAAAGAAGCAGCGYGCLG	654
Dd	557	agaaaagaaggaqgggygglysgsq-----agrgggaqg-----aaaaaaggaqgggygglylg	606
	RESULT_12		
AAR80168	ID	AAR80168 standard; Protein: 831 AA.	
XX	AAR80168;		
XX	DT	01-MAY-1996 (first entry)	
DE	PWISS1 MASP spider silk protein insert product.		
KM	Spider silK; repeat unit; consensus; minor ampullate silk protein;		
KW	spidroin; MiSP; orb web spider; dragline.		
XX	Nephila clavipes.		
OS			
FH	Key Location/Qualifiers		
FT	Misc-difference 272..285		
FM	/note= "represented as indeterminate, corresponds to a highly compressed GC rich region in AAg98470 which could not be sequenced"		
FN	MO9525165-A1.		
PD	21-SEP-1995.		
PF	14-MAR-1995; 95WO-US03139.		
PR	14-MAR-1994; 94US-O209747.		
PA	(UYWY-) UNIV WYOMING.		
PI	Colegin M, Lewis RV:		
DR	WPL: 1995-336970/43. N-PSTDB: AAO98470.		
PT	Polyptide(s) comprising repeated unit amino acid sequences, also cDNAs - derived from minor ampullate spider silk proteins and used to form spider silk fibres		
PS	Claim 2; Fig 1; 86pp; English.		
CC	This sequence represents the product of CDNA clone, pMISS1, encoding the orb web spider minor ampullate silk protein MissP1 has been identified and sequenced. Repeat unit peptides of MissP1 may form part of a larger polypeptide with an amino terminus (see AAR80184-85) and a carboxy terminus (see AAR80186-89). Repeat unit peptides make up spider silk proteins (spidrolins) which in turn aggregate to form the silk fibres. Spider silk fibres have high tensile strength and significant elasticity. An isolated cDNA clone of a silk protein encoding sequence is of use to produce the protein at high yields using recombinant DNA technology.		
XQ	Sequence 831 AA;		

OY	21	GGGCGAAGAAAAAAGAGAGCGGCGTGLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCG	80
Db	79	gggyag--agavaaagaagaggyg-----Tgaggygggygggyagagaaaa--gagag	129
OY	81	GYGGLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCGGCG-----GLGSGCTSGRCGLG	135
Db	130	gaggy-----gTgagagagagaagagagagagaggygggygTgagagagaaaaagagggag	183
OY	136	GCGAGAAAAAAGAGAGAGCGGCGTGLGSGCTSGCPGCGPCQOTSGRCGLGCGGAGAAAA	195
Db	184	gyTgagagagaaaa--gagaggygggyggagagagaaaaaagagsgaggyTgagagag	241
OY	196	AAAAAGAGCGGCGTGLGSGCTSGRCGLGCGGAGAAAAA-----AAAAAG	240
Db	242	gaagagagagaggy-----gggygagagagaaaaaaxxxxxxxxxxxgagagag	293
OY	241	AGCGGCG-----GLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCGGCG-----	286
Db	294	gggygggygagagagaaaaaagagagaggyTgagagagaaaaagagagaggygggygagag	353
OY	287	--LGSGCTSGRCGLG--GCGAGAAAAAAGAGAGCGGCGTGLGSGCTSGCPGCGPCQ	340
Db	354	aaaaagagagaggyTgagagagagaaaaagagaaagagaggygggyg-agagagaaaa	412
OY	341	QTSRCGLGCGGAGAAAAAAGAGAGCGGCGTGLGSGC-----TSRCGLGCG	389
Db	413	agagagaggyTgagagagagaaa--gagaggygggyggagagagaaaaaTgagagag	470
OY	390	GAGAAAAAAGAGAGAGCGGCGTGLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCGGCG	449
Db	471	gTgagagagagaaagagagTg-----agyygggyggagagagagaaaaagagagaggy	521
OY	450	-GLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCGGCG--LGSGCTSGPG-GYCPGQ	504
Db	522	rgag-----agagaaagagagaaagagaggygggyggagagagaaaaagagagaggy	577
OY	505	QTSRCGLGCGGAGAAAAAAGAGAGCGGCG--GLG-----QCTSGRCGLG-GCGA	555
Db	578	srTgagagagagagaaagagagaggygggyggagagagaaaaagagaggyTgagag	637
OY	556	GAAAAAAGAGAGAGCGGCGTGLGSGCTSGRCGLGCGGAGAAAAAAGAGAGCGGCGL	615
Db	638	gaaagagagaaagagaggy-----gggygggygagagaaaaaga--gagTggyg--	682
OY	616	GSQGTSGRCGLGCGGAGAAAAAAGAGAGAGCGGCG	651
Db	683	rgagaggygggygTgagagagagaaaaa--gagaggyg	716
RESULT 13			
AAAY40099			
ID	AAAY40099 standard; protein: 615 AA.		
AC	AAAY40099;		
XX	19-NOV-1999 (first entry)		
DT			
XX	Spider silk protein spidroline minor 1.		
DE			
XX	Spider silk protein; spidroline minor 1; cosmetic; make-up;		
KW	dermatological compositions; hair care; skin care; sunscreen;		
KW	hormone; moisturizer; skin disorder; skin disorder.		
XX	Nephila clavipes.		
OS			
XX	FR2774588-A1.		
PN	13-AUG-1999.		
XX			
PD	11-FEB-1998; 98FR-0001614.		
XX			
PR	11-FEB-1998; 98FR-0001614.		

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:27 ; Search time 48.42 Seconds

(Without alignments)
283.323 Million cell updates/sec

Title: US-09-490-291-4

Perfect score: 3493
Sequence: 1 MRGSHHHHSGMSAGRGCL.....GCGPGQOTSGIRPAKLIN 681Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summariesDatabase : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	718	1	US-08-425-069-2
2	2135	61.1	718	2	US-08-317-844B-2
3	2135	61.1	747	4	US-09-034-177-3
4	1405	40.2	832	1	US-08-209-747-2
5	1405	40.2	832	1	US-08-458-298-2
6	1233.5	35.3	1177	1	US-07-609-716-31
7	1233.5	35.3	1177	1	US-08-175-155-29
8	1233.5	35.3	1177	1	US-08-477-509B-64
9	1233.5	35.3	1177	2	US-08-707-237A-35
10	1233.5	35.3	1177	3	US-08-482-085B-64
11	1233.5	35.3	1177	4	US-08-475-411A-31
12	1233.5	35.3	1177	4	US-08-478-029A-31
13	1176.5	33.7	1059	1	US-08-175-155-48
14	1176.5	33.7	1059	2	US-08-707-237A-54
15	1176.5	33.7	1101	1	US-08-477-509B-83
16	1176.5	33.7	1101	3	US-08-482-085B-83
17	1157	33.1	738	3	US-08-864-038A-3
18	1157	33.1	595	1	US-08-425-069-4
19	1132	32.4	595	2	US-08-317-844B-4
20	1131	32.4	1332	1	US-07-609-716-41
21	1131	32.4	1332	4	US-08-475-411A-41
22	1131	32.4	1332	4	US-08-478-029A-41
23	1129	32.3	1038	1	US-07-609-716-36
24	1129	32.3	1038	4	US-08-475-411A-36
25	1129	32.3	1038	4	US-08-478-029A-36
26	1127	32.3	979	1	US-08-477-509B-89
27	1127	32.3	979	3	US-08-482-085B-89

28	1127	32.3	1050	1	US-08-175-155-54	Sequence 54, App1
29	1103.5	31.6	1018	1	US-08-089-862-11	Sequence 11, App1
30	1103.5	31.6	1018	1	US-08-587-333-18	Sequence 18, App1
31	1103.5	31.6	1018	5	PCT-US94-07776-16	Sequence 16, App1
32	1099	31.5	907	2	US-09-010-928B-4	Sequence 4, App1
33	1095.5	31.4	945	1	US-08-089-862-6	Sequence 6, App1
34	1095.5	31.4	945	1	US-08-587-333-13	Sequence 13, App1
35	1095.5	31.4	945	5	PCT-US94-07776-11	Sequence 11, App1
36	1094	31.3	2100	1	US-08-477-509B-80	Sequence 80, App1
37	1094	31.3	2100	3	US-08-482-085B-80	Sequence 80, App1
38	1094	31.3	2107	1	US-08-175-155-45	Sequence 45, App1
39	1094	31.3	2107	2	US-08-707-237A-51	Sequence 51, App1
40	1083	31.0	745	2	US-09-010-928B-28	Sequence 28, App1
41	1082	31.0	766	1	US-08-175-155-53	Sequence 53, App1
42	1082	31.0	766	1	US-08-477-509B-88	Sequence 88, App1
43	1082	31.0	766	2	US-08-707-237A-61	Sequence 61, App1
44	1082	31.0	766	3	US-08-482-085B-88	Sequence 88, App1
45	1080.5	30.9	870	2	US-09-010-928B-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himman, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 61.1%; Score 2135; DB 1; Length 718;
Best Local Similarity 68.3%; Pred. No. 7.4e-151;
Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;
QY 23 QGAGMAAAGAGAGGCGTGGTGGT--SGRGGLGGGAG-AAAAAAAAAGCAGC 79

Db 588 QGAG-----AAAAAGAGAGGCGTGGTGGGCGT-----AAAAAGAGAGG 636
 QY 651 GGLGS 655
 Db 637 GGVGS 641

RESULT 3

US-09-034-177-3
 ; Sequence 3, Application US/09034177
 ; Patent No. 6127146

GENERAL INFORMATION:

APPLICANT: Lal, Preeti
 APPLICANT: Guegler, Karl J.
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/034,177
 FILING DATE: HEREWITH

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0486 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166
 TELEX:

INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:

LENGTH: 747 amino acids
 TYPE: amino acid

STRANDEDNESS: single
 TOPOLOGY: linear

IMMEDIATE SOURCE:
 LIBRARY: GenBank

CLONE: GI 1174414
 US-09-034-177-3

Query Match 61.1%; Score 2135; DB 4; Length 747;
 Best Local Similarity 68.3%; Pred. No. 7, 6e-151;

Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

QY 23 QGAGAAAAAGAGAGCGTGGTGGGCGT--SGRGGLGCGAG-AAAAAAAAAGAGAG 79
 Db 1 QGAG-----AAAAAGAGAGCGTGGTGGGCGT-----AAAAAGAGAGAG 56
 QY 80 GGVGGLSGTSGRGGLGCGAGAAAAAGAGAGCGTGGTGGGCGTSGRGGLGCGAG 139
 Db 57 GGVGGLSGTSGRGGLGCGAGAAAAAGAGAGCGTGGTGGGCGTSGRGGLGCGAG 107
 QY 140 GAAAAAGAGAGCGTGGTGGGCGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAG 199
 Db 108 G-----AAAAAGAGAGCGTGGTGGGCGT-----AGRGGLGCGAG 142

QY 200 AGAGAGCGTGGTGGGCGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAG 250
 Db 143 AGAGAGCGTGGTGGGCGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAG 201
 QY 251 SGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
 Db 202 SGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 260
 QY 290 QGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 348
 Db 261 QG-----AGRGGLGCGAG-----AAAAAGAGAGAGAGAGAGAGAGAGAGAG 311
 QY 349 GGGAG 390
 Db 312 GGGAG 370
 QY 391 AGAAAAAG 450
 Db 371 AG-----AVAAAAAG 419
 QY 451 LGSQGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
 Db 420 LGNGG-AGRGGLGCGAG-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
 QY 498 GGYGP-GQGTSGRGGLGCGAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 550
 Db 476 GGYGGLSGTSGRGGLGCGAG-----GGGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
 QY 551 GGGAG 590
 Db 532 GGGAG 587
 QY 591 QGAGAAAAAG 650
 Db 588 QGAG-----AAAAAG 636
 QY 651 GGLGS 655
 Db 637 GGVGS 641

RESULT 4

US-08-209-747-2
 ; Sequence 2, Application US/08209747

Patent No. 5733771

GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph V.

TITLE OF INVENTION: CDNA's Encoding Minor Ampullate Spider
 NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747
 CITY: Falls Church

STATE: Virginia
 COUNTRY: USA

ZIP: 22040-3487

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/209,747
 FILING DATE: 14-MAR-1994

CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:

NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1447-104P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-205-8000

COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-29

Query Match 35.3%; Score 1233.5; DB 1; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;

Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

13 MASGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GGAGAAAAA 69
32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91
70 AAAAAAGAGCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAGAGAGAGAG 128
92 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 151
129 SGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---SGRGLGGG 185
152 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA 211
186 GCGAGAAAAAAGAGAGGCGGCGGCGTSGRGLGGGAGAAAAAAGAGAG 242
212 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 271
243 GCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 295
272 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 331
296 GGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---SGRGLGGG 352
332 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 391
353 AGAAAAAAGAGAGGCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAG 409
392 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 451
410 YGGLGSGCGTSGRGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAG 462
452 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 511
463 GGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---SGRGLGGGAG 519
512 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 571
530 AAAAAAAGAGAGGCGGCGGCGTSGRGLGGGAGAAAAAAGAGAGAG 576
572 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 631
577 LGSQGTSGRGLGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 629

DB 632 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGAGAGAGAGAGAGAGAG 691
OY 630 GAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GGAGAAAAA 667
DB 692 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGAGAG 729

RESULT 8
US-08-477-509B-64

Sequence 64, Application US/08477509B

Patent No. 5770697

GENERAL INFORMATION:

APPLICANT: Ferrari, Franco A

APPLICANT: Cappello, Joseph

APPLICANT: Crissman, John W

APPLICANT: Dorman, Mary A

TITLE OF INVENTION: No. 5770697e1 Peptides Comprising Repetitive

TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same

NUMBER OF SEQUENCES: 112

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hobach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentlin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,509B

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/175,155

FILING DATE: 29-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/053,049

FILING DATE: 22-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/114,618

FILING DATE: 29-OCT-1987

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 06/927,258

FILING DATE: 04-NOV-1986

ATTORNEY/AGENT INFORMATION:

NAME: Treccarlin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989

TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 64:

SEQUENCE CHARACTERISTICS:

LENGTH: 1177 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-477-509B-64

Query Match 35.3%; Score 1233.5; DB 1; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;

Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

13 MASGRGLGGGAGAAAAAAGAGAGGCGGCGGCGTSGRGLG---GGAGAAAAA 69
32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91

```

QY 70 AAAAGAGAGGCGYGLGSGCTSGRGLGCGGAGAAAAAAG-GAGGCGYGLGSGCT 128
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 92 GAGSGAGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151
QY 129 SGRGGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 152 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211
QY 186 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 212 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
QY 243 GCGYGLGSGCTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 272 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331
QY 296 GGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 332 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 353 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 392 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451
QY 410 YGGLGSGCTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 452 GAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
QY 463 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 512 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 571
QY 520 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 576
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 572 GSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
QY 577 LGSGCTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 629
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 632 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691
QY 630 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 692 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729

RESULT 9
US-08-707-237A-35
: Sequence 35, Application US/08707237A
: Patent No. 5830713
: GENERAL INFORMATION:
: APPLICANT: Ferrari, Franco A.
: APPLICANT: Capello, Joseph
: APPLICANT: Crisman, John W.
: APPLICANT: Dorman, Mary A.
: TITLE OF INVENTION: METHODS FOR PREPARING SYNTHETIC
: TITLE OF INVENTION: REPEITIVE DNA
: NUMBER OF SEQUENCES: 108
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
: STREET: Four Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/707,237A
: FILING DATE: 03-SEP-1996
: CLASSIFICATION: 435

```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/175,155
: FILING DATE: 29-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/053,049
: FILING DATE: 22-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/609,716
: FILING DATE: 06-NOV-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/269,429
: FILING DATE: 09-NOV-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/114,618
: FILING DATE: 29-OCT-1987
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 06/927,258
: FILING DATE: 04-NOV-1986
: ATTORNEY/AGENT INFORMATION:
: NAME: Treccartin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: A-55186-10/MHD
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 398-3249
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEO ID NO: 35:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1177 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-707-237A-35

Query Match 35.3%; Score 1233.5; DB 2; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.3e-84;
Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

QY 13 MASRGGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 32 MGAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91
QY 70 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 128
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 92 GAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151
QY 129 SGRGGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 185
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 152 AGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211
QY 186 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 242
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 212 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
QY 243 GCGYGLGSGCTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 295
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 272 SGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331
QY 296 GGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 332 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 353 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 409
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 392 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451
QY 410 YGGLGSGCTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
| : | | | | | : | | | | | : | | | | | : | | | | | : | |
Db 452 GAGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
QY 463 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 519

```

[illegible]

US-08-482--085B-64

Query Match 35.3%; Score 1233.5; DB 3; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3.5e-84;
Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps

Dy 13 MASRGGTGTGCGAGCAAAAAAAAAAAGAGAGGCGGTGGTGCTSGRGTLG--GCGAGAAAA 69
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 32 MGASGAGAGSAGA 91

Oy 70 AAAAAAGAGAGCGGTGTCGTSQGTSSRGGLGCGAGCAAAAAAAAAAAAG-CAGCGGTGTCGTT 128
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 92 GAGGAGAGAGSAGA 151
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 129 SGRGRLGCGCGCAAAAAAAAAAAGAGAGCGGTGGTGCTSGTSPPGCTGPPOQT---SGRGLG 185
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 152 AGSGA 211
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 186 GCGAGCAAAAAAAAAAAGAGAGCGGTGTCGTSQGTSSRG--GLGCGAGCAAAAAAAAAAGAG 242
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 212 GSGA 271
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 243 CGGTGTGCTSGTSRGGTGGGGAFAAAAAAAAAAAGAGAGCGGTG-----GLGSGTSGR 295
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 272 SGA 331
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 296 GGLGCGAGCAAAAAAAAAAAGAGAGCGGTGGTGCTSGTSPPGCTGPPOQT---SGRGLGCGG 352
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 332 GA 391
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 353 AGCAAAAAAAAAAGAGAGCGGTGTCGTSQGTSSRG--GLGCGAGCAAAAAAAAAAAGAGAGCGG 409
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 392 AGA 451
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 410 YGGLGCTSGTSRGGTGGGGAFAAAAAAAAAAAGAGAGCGGTG-----GLGSGTSGRGL 462
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 452 GAGSAGA 511
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 463 GCGCAGCAAAAAAAAAAAGAGAGCGGTGTCGTSQGTSSPCTGTPPOQT---SGRGLGCGAGAGA 519
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 512 AGSGA 571
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 520 AAAAAAAAAAGAGAGCGGTGGTGCTSGTSRG--GLGCGAGCAAAAAAAAAAAGAGAGCGGTG 576
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 572 GSGA 631
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 577 LGSQGTSSRGGTGGGCAAAAAAAAAAAGAGAGCGGTG-----GLGSGTSSRGGTGGG 629
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 632 SGA 691
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

Oy 630 GAGCAAAAAAAAAAAGAGAGCGGTGGTGCTSGTSPPGCTGP 667
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 692 GA 729

RESULT 11
US-08-475-411A-31
Sequence 31, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111


```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-31

```

```

Query Match      35.3%; Score 1233.5; DB 4; Length 1177;
Best Local Similarity 41.0%; Pred. No. 3,5e-84;
Matches 286; Conservative 58; Mismatches 311; Indels 43; Gaps 11;

```

```

QY 13 MASGRGLGGGAGAAAAAAGAGGCGGGLGSGGTSGRGGLG---GCGAGAAAAA 69
DB 32 MGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 91
QY 70 AAAAGAGAGCGGCGGLGSGGTSGRGGLGCGAGAAAAAAGAGAGAGGCGGGLGSGGT 128
DB 92 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 151
QY 129 SGRGGLGCGGAGAAAAAAGAGAGCGGCGGGLGSGGTSGRGGLG---SGRGGLG 185
DB 152 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 211
QY 186 GCGAGAAAAAAGAGAGCGGCGGGLGSGGTSGRG---GAGSGAGAAAAAAGAGAG 242
DB 212 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 271
QY 243 GCGGGLGSGGTSGRGGLGCGAGAAAAAAGAGAGCGGCGG---GLGSGGTSGR 295
DB 272 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 331
QY 296 GCGGAGAGAGAAAAAAGAGAGCGGCGGGLGSGGTSGRGGLG---SGRGGLG 352
DB 332 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 391
QY 353 AGAAAAAAGAGAGCGGCGGGLGSGGTSGRG---GAGSGAGAAAAAAGAGAGCGG 409
DB 392 AGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 451
QY 410 YGGLGSGGTSGRGGLGCGAGAAAAAAGAGAGCGGCG---GLGSGGTSGRGGL 462
DB 452 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAG 511

```

```

QY 463 GCGAGAAAAAAGAGAGCGGCGGGLGSGGTSGRGGLG---SGRGGLGCGAG 519
DB 512 AGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 571
QY 520 AAAAAGAGAGAGCGGCGGGLGSGGTSGRG---GLGCGAGAAAAAAGAGAGCGG 576
DB 572 GCGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 631
QY 577 LGSGGTSGRGGLGCGAGAAAAAAGAGAGCGGCG---GLGSGGTSGRGGLG 629
DB 632 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGA 691
QY 630 GAGAAAAAAGAGAGCGGCGGGLGSGGTSGRGGLG 667
DB 692 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 729

```

```

RESULT 12
US-08-478-029A-31
; Sequence 31, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: Functional Recombinantly Prepared
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-029A-31

```



```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,509B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1101 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-509B-83

```

```

Query Match      33.7%; Score 1176.5; DB 1; Length 1101;
Best Local Similarity 40.8%; Pred. No. 5.3e-80;
Matches 270; Conservative 60; Mismatches 319; Indels 13; Gaps 7;

QY 11 GSMASGRGLG-GGAGAGAAAAAGAGGAGGCGYGGGLGSGGTSGRGLGGGAGAAAA 69
DB 432 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 491
QY 70 AAAAAAGAGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAGGCGYGGGLGSGGTS 129
DB 492 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 551
QY 130 GRGGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGTSGPG-GYGPQQTSGRGLGGG 188
DB 552 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 608
QY 189 AGAAAAAGAGGAGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAGGAGGCGYGG 248
DB 609 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 668
QY 249 LGSQGTSGRGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGTSGRGLGGGAGAAA 308
DB 669 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 728
QY 309 AAAAAAGAGGCGYGGGLGSGGTSGPG-GYGPQQTSGRGLGGGAGAAAAAGAG 367
DB 729 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 785
QY 368 GCGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGTSGR 427
DB 786 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 845
QY 428 GAGAAAAAGAGGAGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAGGAGGCGY 487
DB 846 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 905
QY 488 GLGSGGTSGPG-GYGPQQTSGRGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGT 546
DB 906 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 962

```

```

QY 547 RGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAG 606
DB 963 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 1022
QY 607 AGGCGYGGGLGSGGTSGRGLGGGAGAAAAAGAGGAGGCGYGGGLGSGGTSG 666
DB 1023 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 1082
QY 667 GQ 668
DB 1083 GR 1084

```

Search completed: July 3, 2001, 14:56:34
Job time: 524 sec

This Page Blank (uspto)

Query Match 33.6%: Score 1173; DB 2; Length 1660;
 Best Local Similarity 35.7%: Pred. No. 5,4e-58;
 Matches 331; Conservative 36; Mismatches 281; Indels 280; Gaps 37;

11 GMSMSGRGGLGCGGAGAAAAAAGAG-----GCGGYG-----46
 451 GPVSGGNGCKKGNHAYVAGCHGNGAGNGGLVGGGAGGCHGDPAAGAYADMTAI 510
 47 -LGSGT-----SGRGLGCG--GAGAAAAAAGAGAGGCGGGLGSG-----88
 511 FLGSGGTGEGEDGNGAGAGAGAGAHGDCAGAGAGGAGGAGGAGHFNATVSD 570
 89 -----GTSRGLGCGGAGAAAAAAGAGAGGCGGGL-----123
 571 GGNGGCGAGGRCGPDGAG--GAGGDPAGRAGSGGVDGAGAGAGAPNGSGSGRDM 629
 124 -----GSGGTSGRGLGCGA-----GAAAAAA 147
 630 AFRKDDGAGDGDGPPGAGGKAGAGACATGATGATVHSGNGKCGNADATVAG 689
 148 AAGAGAGCGYGL--GSGGTSGPGY-----PG---QQTSGRG 182
 690 ANGCGAGAGNGGLVDCAGAGDGGGAAAGANVGEDADGTLGCPGEGSEANGCG 749
 183 GLGCGAGAAAAAAGAGAGGCGGGLGSGGTSGRGLGCGGAGAAAAAAGAG 242
 750 GYGCGGAGAGCGDGGAGSALGSGNGRGDAQAG---GAGGAGAGAGAGSVSDGG 805
 243 GCGYGLSGGTSGRGLGCGGAGAAAAAAGAG--GCGYGLGSGGTSGRGLGCG 301
 806 PGCKGAGAGAGAGAGAG--GCGGKAGSAGDSAPAVGAGAGCGKGDGIVG---GYGDDGPPGD 862
 302 G--AGAAAAAAGAGAGGCGYGL--GSGGTSGPGY-----PGQQTSGRG 347
 863 GGAAGAAPAGYVSHGVTGCGDGLGAGAGNGGSDGCGDGDGDPG--AGGLG 920
 348 LCG--GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 385
 921 LGDGGNGTRAAISVDASDHGPGSGGNGNGNGAGASVAGAGAGNGDGGAGRVGDDG 980
 386 LGGGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
 981 AGGNGGDAAGANGANGSAPGSDALALGPPGNGCGGDAAGAGAGAGAGAGAG 1040
 436 AAAAAAGAGAGGCGYGLGSGGTSGR-----GGLGCGAGAAAAAAGAGAGG 486
 1041 DCGAGGAG 1100
 487 GG-----LGSQGTSGPGY-----PGQQTSG 507
 1101 GGAAPSGTVGSHGTGGVGGDGLGAGAGVAGAGNGGIGITVGAAGAGAGAGAG 1158
 508 GRGGLGCG-----GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 543
 1159 GNGGGLGCGDGNKTSANVDAKHPPLTNGDGGVGNNGAKAAAAAGDGGDGGD--GNGG 1216
 544 TSGRGLGCGGAGAAAAAAGAG--GCGGYG-----LGSQGTSGRGLGCGG 593
 1217 LFGDGGAGCGDGTAAELAGDGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
 594 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644
 1277 VAGGAG 1336
 645 AGGAGYGLGSGGTSGPGYGPQQTSG 672
 1337 NGCGGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364

RESULT 9
 HT0846
 hypothetical glycine-rich protein RV3345c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: H70846
 R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garler, T.; Churcher, C.; Harris, D.; Gordon,
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A: Reference number: A70500; MUID:98295987
 A: Accession: H70846
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-1538 <COL>
 A: Cross-references: GB:AL021841; GB:AL123456; NID:g3261517; PID:CAI17117.1; PID:g326
 A: Experimental source: strain H37RV
 C: Genetics:
 A: Gene: RV3345c
 C: Superfamily: collagen alpha 1(IV) chain

Query Match 33.1%: Score 1157; DB 2; Length 1538;
 Best Local Similarity 34.4%: Pred. No. 4e-58;
 Matches 317; Conservative 36; Mismatches 308; Indels 260; Gaps 31;

1 MRGSHHHHSGMSAGRGGLGCGGAGAAAAAAGS-----AGCGYGLGSGGTS 53
 549 LNSGAGANGTAPTSCKNGNGAGCATPTVAGENGAGGNGHGSYVNGAGAGAGNVA 608
 54 G-----RGGLGCGA--GAAAAAAGAGAGCGYGLGSG-----GTS 91
 609 GTGLALNGGNGNGIGGNGSAACTGGDGGKNGCAGANGDPESASANGANGGCGNG 668
 92 GRGGLGCGGAGAAAAAAGAGAGGCGYGLGSGGTSGRGLGCGGAG-----139
 669 GNGGIGGKGAFATFAAAGAGAGAGNG---GNVAVAGCGAGGAGKAIIPAKKATGADG 724
 140 -----GAAAAAAGAGAGAGCGYGL--LGSQGTSGPGYGPQQT-----T 178
 725 TAPTSCKGCGNGNGNASTPVVAGNGGDDGKSGGNGNGNGAGAGAGAGAGAG 784
 179 SGRGLGCGGAGAAAAAAGAG--AGCGYGLGSGGTSGRGLGCGGAGAAAAA 236
 785 SDDSGTSGTDGAGGNGAGAGGAGGTLAGHNGNGK--GNGGCGGIGGAGERGADGAPN 842
 237 AAGAGCGGCGYGLGSGGTSGRGLGCG--GAGAAAAAAGAGAGAGCGYGLG 286
 843 ANGANGENG---GSGGNGCGDAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
 287 -----LGSQGTSGRGLGCG-----301
 899 DCGAGENGLNSGAMLPGGGTGYNPPTGGNGGNGAGAGCGKAGTGSLTGLDGTDT 958
 302 -----GAGAAAAAAGAGAG-----AGCGYGL-----LGSQGTSG 331
 959 PNGGNGNGNGCKGCTAGNGSAGAGAGNGGSGSLNGDAGNGGNGGALNQAFEFPG 1018
 332 GPGGY-----GPGQQTSGRGLGCGGAGAAAAAAGAGAGCGYGLGSGGTSG 387
 1019 GGGGNGGNGGAGAMINGGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
 388 GCGAGAAAAAAG 436
 1079 GPGGAGDPRASGVGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
 437 AAAAG-----AGCGYGLGSGGTSGRGLGCGGAGAAAAAAGAGAGAG 484
 1139 NGGLGNGGVSETFGAGAGAGNGYGP--GPEGNGGLGNGAGAGAGAGAGAGAG 1196
 485 GYGGLGSGGTSGPGYGPQQTSGRGLGCGGAGAAAAAAGAGAGAGAGAGAGAG 532
 1197 GKGNGCGD--GNVGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
 533 GCGYGLGSGGTSGRGLGCGGAGAAAAAAG-----AAGAGAGCGYGLGSGGTS--GRGL 588

Matches 303; Conservative 18; Mismatches 309; Indels 135; Gaps 32:

```
Oy      16 GRGGLGGAGCAAAAAAAAAAACGACGGYGGLGSG-----GTSGR   55  
         ||| | | | | | | | | | | | | | |  
Db       3 GNGGAGGSAPCATIGAGGPAGLIGVGAGGASVAVGYCGAAGAALEFAGAGA   62  
  
Oy      56 GGLGGGAGAAAAAAAACA-----GGCGTGLSOGTSGRGTLGGCAGAAAAAA   108  
         :| | | | | | | | | | | | | | |:  
Db       63 GGAGSSG--GSGMAAGACAGAGAGGLEPASGSGCFPGPSTGTGCAGCTGAAGLFASGV   121  
  
Oy     109 AAAAGAGGCGGCGTSGRGTLGGGAGAAAAAAAAGAGGGCYGL- GSOGTS   167  
         |||| | | | | | | | | | | | | | |  
Db     122 GCTGGAGASSGCTGGG--GTGAGGLFPASGAGAGSGSGCTGAGCTGAGGLEAGAG   179  
  
Oy     168 GPCCGPPCOQTSGRCGLGSO-----GAGAAAAAAAAGAGA--GGGYGGL--GSOG   215  
         ||| | | | | | | | | | | | | | |  
Db     180 GLG--GQGNHTTGCHGAGGASAGLLALDPDGAAGAGAAATTGTGAGAGAKRAGLLFGSGG   237  
  
Oy     216 TSGRGLGGO-----GACAAAAAAAAGAGACGGYGLGISOGTSGRGLGGGAGA   267  
         ||| | | | | | | | | | | | | | |  
Db     238 AGSGCAAGTTPEDTNSGAGAGAGCACGLLPSCGAGSGSGAGPANGSTGAGAGAGGA   297  
  
Oy     268 -----AAAAAAAAGAGGCGGYGGLGSG-----GTSGRGLGGGAGAAAAAAA   314  
  
Db     298 CLINGMGKSGSGTSVATAGAGNKGAGAGAGGLIGNMNGSGGMG-----DAP   348  
  
Oy     315 CGACGGCGYG--LGSQTSGBPQYVP-----GQOTSGRGLGGQG   352  
         ||| | | | | | | | | | | | | | |  
Db     349 GSTGVTGTLGLLDGANNAPOSTNP LHTAQQAALAANAIPTAYTRPLIGNGANARP   408  
  
Oy     353 AAAAAAAAAAAAGAGCGYGLGISOGTSGRGLG--GAGAAAAAAAAGAGAGCGY   410  
         :| | | | | | | | | | | | | | |:  
Db     409 SGAPCGHGWFLFCGGCTGGSGVS--GAGCDGAGGILFFGAGAGAGAGAAVTGTATGGS   466  
  
Oy     411 GGLGSO----GTSGRGLG--GAGAAAAAAAAGAGAG---GCGYGLSOGTSGRG   461  
  
Db     467 GGAGGALLFFGAGAGAGSSSIGTFPAAGAGGPBGAGLFNKGAGAGAGSGYS--GG   524  
  
Oy     462 LGCGAGCAAAAAAAAGAGAGCGYGGYGLSOGTSPGCGPQOTSGRGLGGGAGAAA   521  
         ||| | | | | | | | | | | | | | |  
Db     525 AGEG-GAGAGAGLFAAGGAGAGAGCSGNWVGAGAGAGVG---GLFGAGAGAGSSGGSV   580  
  
Oy     522 AAAAAAAGAG-----GCGYGLSOGTSGRGLGGGAGAAAAAAAAGAGAGCGY-   574  
         ||| | | | | | | | | | | | | | |  
Db     581 AGDSGAGNAAGLAPLAGAGAGCGGCGFD-TGAGAGPGBDAAGLLVSGYVAGAGFGLT   639  
  
Oy     575 -GGLSOGTSGRGGL--GGGAGAAAAAAAAGAGA-CGGYGL- GSOGTSGRRLGGQ   629  
         :| | | | | | | | | | | | | | |:  
Db     640 TGCPGAG--GPAGLIFSGGAGGAGRTDLFGAGAGGAFMLIGNMNGAGAGGN   697  
  
Oy     630 GAGAAAAAAAAGAGAGCGYGLSOGTS--GPCGYGPQQTS   672  
         || | | | | | | | | | | | | | |  
Db     698 GCGDGPCGAAF--GLNGMNGMNGCTGTAGSPAGAGGSLIG   740
```

RESULT 14
H70987
hypothetical glycine-rich protein RVJ1759c - Mycobacterium tuberculosis (strain H37RV)
N:Alternate names: wag22 antigen homolog
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
X:Accession: H70987
R:Colo, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Majumder, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellom, S.; Squares, S.
Neure 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: AF0500; MUID:98295987
A:Accession: H70987
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-914 <COL>

A:Cross-references: GB:255890; GB:AL123456; NID:q3242245; PIDN:CAMB09322.1; PID:q21310

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV1759c; wag22

C:Superfamily: elastin

```
Query Match          31.7%, Score 1106.5; DB 2; Length 914;
Best Local Similarity 38.2%; Pred. No. 1.8e-55;
Matches 306; Conservative 31; Mismatches 304; Indels 159; Gaps 32;

QY 11 GSMAAGRGGLGGGAGCAAAAAAAGAG----QGYGGL-GSQGTSGRGLGGGAGA 65
Db 116 GNGANGAPGTGNGDDACMLIGNGAGSGGAKANGAGGPEGAAGLFNGAGGAGSTA 175
QY 66 AAAAAAAAAA-----GQGYGLGSQGTSG--GRGGLGGGAGAAAAAAGCA 115
Db 176 TANNIGGAGAGGAGSAMLFGAGGAGAGGAATSLVGTGTCGTGNNMGLAAGAGGA 235
QY 116 G-----QGYGGLGSQ-----TSGRGGLGGG-----AGAAAAAAGG-AGQGYG 159
Db 236 GGFSTAGAGAGAGAGAGLFTTGGVGGAGGCHTGGAGAGAGAGLFGAGGAGGCGFG 295
QY 160 GLGSQGTSGPGYGRGQQTSGRGLGGGAGCAAAAAAAG-----AGQGYG 209
Db 296 DHGTGTGTGAGGDDGGGGLFGAGGAGGAGSGLTTGGAAGNGNAGTSLGAGAGGTTG 355
QY 210 GLGSQGTSGRGLGGGAGCAAAAAAAGAG-----QGYGGLGS-----QGTSG 256
Db 356 GAG--GYTGGGKGGAGAGGAGNAGMLFGSGGGGTGCTGFAAGGGGYGGSAGMLSGSG 413
QY 257 RGLGGGQ--AGAAAAAAGAGAGGAGGGLGSQ--TSGRGGLGGGAGCA-----A 306
Db 414 SGGAGGSGGPACTAAGAGAGAGAPGLIGNGNGNGSGGTGCVGAGGNAVLIGNG 473
QY 307 AAAAAAGGAGCGGYGGL-----325
Db 474 EGGIGLALGKSGFGFGGILLGADYNAPESTSPWHNLQODILSFINEPTALTGRPLIG 533
QY 326 -GSQGTSGPG--GYGQQQTSGRGLGGGAGCAAAAAAAGCA-----GQGYGGLSQ 378
Db 534 NCDSTPGTGDGAGAGMLFNGNGAGAGA-AGTNGSAGAGAGGILLFTGTGAGAGAGV 592
QY 379 GTSGRGLGGGAGCAAAAAAAGAGAGGAGGAGGGLGSQGTSGRGLGGGAGCAAAAAA 438
Db 593 GTAGAGAGAGAGAGSFL-----ISGCTGCVGAATT--TGVGAGGAGNALLIGAA 642
QY 439 AAGAGAGGY--GGLSQGTSGRGLG-----QAGAAAAAAGAGAGGAGGGLGSQ- 492
Db 643 GLGCGCGGCAFTGYTTGAGGTGGAAGLFPANGAGAGAGAGTGSTAGAGAGAGGLYAHG 702
QY 493 GTSRPGCYPRGQQTSGRGLGGGAGCAAAAAAAGAG--QGYGGLGSQGTSG----- 545
Db 703 GTGCGPGNNGSTGAGGTGAGGP--GLYAGSGGAGGGHAGAGGGGAGGAGSLTTANA 761
QY 546 --GRGGLGQ-----GAGAAAAAAGAGAGGAGGAGGGLGSQGTSGRGLGGGQ----- 592
Db 762 SGGAGGSGGSSLSKRAAGAGAGAGSAGLFTGSGAGAGNGYSLNGTCDP6GTSGAQDINGL 821
QY 593 ----AGAAAAAAGAGAGGAGGAGG--LGSQGTSGRGLGGGAGCAAAAAAAGAGAGQ 647
Db 822 RSGFGAGAGAGASPTGAGAGNGAGGAGKAGLTYNGGDDGAGAGGATSGKAGAGNAVYIGN 881
QY 648 GGYGGLGSQ--GTSRPGCYG 665
Db 882 GGNGAGNAGAGGTAGAGAG 901

RESULTS
70812
Hypothetical glycine-rich protein RV0834c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:32 ; Search time 37.35 Seconds

(Without alignments)

624,578 Million cell updates/sec

Title: US-09-490-291-4

Sequence: 1 MRGSHHHHSGMSASGRGL.....GCGPFGQOTSGIRPAKLN 681

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2135	61.1	747	1 SPDL_NEPCL	P19837 nephila cla
2	1277.5	36.6	1901	1 YZ08_MYCTU	O53533 mycobacteri
3	1168	33.4	5263	1 FBOH_BOMO	P05780 bombyx mori
4	1135	32.5	627	1 SPDL_NEPCL	P46804 nephila cla
5	1106.5	31.7	914	1 WA22_MYCTU	O06794 mycobacteri
6	1044.5	29.9	778	1 YQ34_MYCTU	P71933 mycobacteri
7	1028	29.4	957	1 Y278_MYCTU	P56877 mycobacteri
8	988	28.3	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.9	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.9	864	1 ELS_RAT	O99372 rattus norv
11	869	24.9	603	1 YD25_MYCTU	O10637 mycobacteri
12	853.5	24.4	465	1 GRP2_PHAVU	P10496 phaseolus v
13	775.5	22.2	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.8	641	1 EBN1_EBV	P03211 epstein-bar
15	704.5	20.2	750	1 ELS_CHICK	P07916 gallus gall
16	691	19.8	730	1 ELS_HUMAN	P15502 homo sapien
17	686	19.6	498	1 Y118_MYCTU	O50675 mycobacteri
18	679	19.4	1733	1 VNUA_PPKA	P33485 pseudobacile
19	678	19.4	543	1 YP91_MYCTU	O50650 mycobacteri
20	673	19.3	384	1 GRP1_PETHY	P09789 petunia hyb
21	669	19.2	481	1 LORT_MOUSE	P18165 mus musculu
22	665	18.8	1027	1 CAEP_RIFPA	P30754 riftia pach
23	644	18.4	338	1 GRP_ARATH	P27483 arabidopsis
24	636	17.9	515	1 Y140_MYCTU	O50594 mycobacteri
25	635	17.9	1373	1 CA21_MOUSE	O01149 mus musculu
26	631	17.8	1466	1 CA13_HUMAN	P02461 homo sapien
27	616	17.6	434	1 YK96_MYCTU	O10707 mycobacteri
28	608	17.4	1453	1 CA11_CHICK	P02457 gallus gall
29	607.5	17.3	1418	1 CA12_HUMAN	P02458 homo sapien
30	604.5	17.3	1460	1 CA11_CANFA	O93717 bos taurus
31	603.5	17.3	1469	1 CA13_BOVIN	P04258 bos taurus
32	602.5	17.2	463	1 YA68_MYCTU	O53416 mycobacteri
33	602	17.2	1464	1 CA11_HUMAN	P02452 homo sapien

34	598	17.1	1459	1 CA12_MOUSE	P28481 mus musculu
35	597.5	17.1	1464	1 CA13_MOUSE	P08121 mus musculu
36	597	17.1	1366	1 CA21_CANFA	O46392 canis famli
37	596	17.1	1366	1 CA21_HUMAN	P08123 homo sapien
38	595	17.0	1363	1 CA21_BOVIN	P02465 bos taurus
39	594.5	17.0	1453	1 CA11_MOUSE	P11087 mus musculu
40	573.5	16.4	1362	1 CA21_CHICK	P02467 gallus gall
41	563	16.1	1362	1 CA24_CAEEL	P17140 caenorhabd
42	568	16.0	1367	1 ICEN_XANCT	P18127 xanthomonas
43	557.5	16.0	1758	1 CA14_MOUSE	P17139 caenorhabd
44	555	15.9	1775	1 CA14_DROME	P08120 drosophila
45	554.5	15.9	779	1 CA11_BOVIN	P02453 bos taurus

ALIGNMENTS

```

RESULT 1
ID SPDL_NEPCL STANDARD; PRT: 747 AA.
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxId=6915;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
  proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
  Araneus bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
CC - FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
  CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
  PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
  WITH ELASTIC AMORPHOUS SEGMENTS.
CC - SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC - SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37137; AAA29380.1; -
DR EMBL: U03848; AAB60212.1; -
DR PIR: A36068; A36068.
KW Silk; Repeat.
FT DOMAIN 1 1 655 25 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 1 1 25
FT REPEAT 26 38
FT REPEAT 39 66
FT REPEAT 67 96
FT REPEAT 97 130
FT REPEAT 131 158
FT REPEAT 159 191
FT REPEAT 192 204
FT REPEAT 205 235

```



```

OY 296 GGLGGGAGAGAAAAAGAGGAGGGLGSGTSGPGGPGGQGTSGRGLGGGAGA 355
DB 273 G-----SAAAAAAGPAGTG-----PGGPGGPGGPGGQ-----GPGSAGS 310
OY 356 AAAAAAAGAGGAGGAGGGLGSGTSGRG-----GLGSGGAGAAAAAAGAGAG 410
DB 311 AAAAAAAGAGGAGGAGGGLGSGTSGRG-----GLGSGGAGAAAAAAGAGAG 369
OY 411 GGLGSGTSGRGLGGGAGAAAAAAGAGGAGGGLGSGTSG-----RGLGSGGAG 468
DB 370 GP-GGGGPGGPG-----SASAAAAAAGAGGAGGAGGAGGAGGAGGAGGAG 417
OY 469 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 528
DB 418 SASAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 463
OY 529 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588
DB 464 --AGPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 500
OY 589 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 648
DB 501 AGGPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555
OY 649 GYGGLGSG-----GTSGPG 662
DB 556 SSAALSSVISNAVSQIGASMPG 577

```

```

RESULT 5
WA22_MYCTU STANDARD; PRT; 914 AA.
ID WA22_MYCTU
AC 006794;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE WAG32 ANTIGEN PRECURSOR.
GN WAG32 OR RV1759C OR MTC128.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekela F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feldwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon M., Squares S., Squires R., Sulten J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA *Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*;
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGSR
SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 295890; CAB09322.1; -
CC DR HSSP; P41140; 2SFA.
CC DR Tubercultist; RV1759C; -
CC DR Interpro; IPR000084; -

```

```

DR Pfam: PF00934; PE: 1.
KW Antigen; Repeat; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 914 WAG22 ANTIGEN.
SQ SEQUENCE 914 AA; 74354 MW; F6953C3DBEBE6AC8 CRC64;

```

```

Query Match 31.7%; Score 1106.5; DB 1; Length 914;
Best Local Similarity 38.2%; Pred. No. 6.8e-43;
Matches 306; Conservative 31; Mismatches 304; Indels 159; Gaps 32;

```

```

OY 11 GSNAAGRGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 65
DB 116 GNAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 175
OY 66 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 115
DB 176 TANNIGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 235
OY 116 G-----GCGGGLGSG-----TSRGLGAGG-----AGAAAAAAGAG-AGCGG 159
DB 236 GGFSTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 295
OY 160 GLGSGTSGPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 209
DB 296 DHGTLTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 355
OY 210 GLGSGGT-----SGRGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 256
DB 356 GAG--GTFPGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 413
OY 257 RGLGAGG--AGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 306
DB 414 SGAGAGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 473
OY 307 AAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 325
DB 474 EGGTALAGKSGFGGFGGLLGLADGYNAPESTSPMHLODILSFINEPEALTGRLIG 533
OY 326 -GSGTSGPG--GYPGGQGTSGRGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 378
DB 534 NGDSGTPTGTDGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 592
OY 379 GTSRGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 438
DB 593 GTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 642
OY 439 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492
DB 643 GLGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 702
OY 493 GTSRGLGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 545
DB 703 GTGCGPGNGSGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 761
OY 546 --GRGLGAG-----GAGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 592
DB 762 SGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 821
OY 593 ----AGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 647
DB 822 RSGFGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 881
OY 648 GYGGLGSG--GTSRGLGAG-----AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 665
DB 882 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 901

```

```

RESULT 6
YQ34_MYCTU STANDARD; PRT; 778 AA.
ID YQ34_MYCTU
AC P71933;
DT 01-NOV-1997 (Rel. 35, Created)

```


Db 495 SCGAGNATGNGAGGNGMALLGAAGACGAGAGAVG--GNGGAGNGN-GLFANGGAGGPG 551

QY 404 G-----AAGGCGGGLGSG-CTSGRCGTCGCGACAAAAAAGAGAGCGGCGGGLGSGCTSG 458

Db 552 GFGSPAAAGGTCGAGGNGGCTFGAGAGTTGAGAGGSLTLAGACGAGAGNG-----GLFG 601

QY 459 RCGTGCAGGAGAAAAAAGAGAGGCGGGLGSGCTSGPGGCGPQOOTSRRGGLGCGGAC 518

Db 602 AGCTGGAGGSHSTAAGVSGGAGA--GSDADLLSLTAGSGGAGSGGSSLTAAGVVGGIG-- 657

QY 519 AAAAAAAAAAGAGCGGCGGGLGSGCTSGRCGGLGSG-----GAGAAAAAAGAGAGGCGG 573

Db 658 ---GAGCLTFGSGGAGSGGFSNNGNGAGAGAGDAGLLVSGGAGGAGASAT-GAAATGC 713

QY 574 YGGLSGOSTSGRCGLGCGAGAAAAAAGAGAGCGGGLGSGCTSGRCGLGCGGAGA 633

Db 714 DGGAG--GKSGAFLGDDGCGAGCATGSLGAFHIHIGKGGV--GSAVLIGNGNGNGNS 769

QY 634 A-AAAAAAAAAAGAGCGGCGG--LTSOGTSG 660

Db 770 GNAKSGGAPGPGGAGGAGLLLENGLNG 799

RESULT	9	
ELS_MOUSE		
ID	ELS_MOUSE	STANDARD;
		PRT; 860 AA.

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ELASTIN PRECURSOR (TROPOLASTIN).
GN ELN.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;

RC STRAIN=BALB/C; TISSUE=Lung;
RX MEDLINE=95130069; PubMed=7829060;
RA Wydner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;

RT mouse chromosome 5 in a region of linkage conservation with human
RT chromosome 7.⁴
RL Genomics 23:125-131(1994).

CC INTO AN EXTENSIBLE 3D NETWORK.
CC -I- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -I- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/sib-ch>).
CC or send an email to license@sib-sb.ch.

DR	EMBL	U08210; AAA80155.1;	-
DR	HSSP:	P04002; 1WEA.	
DR	MCD:	MGI:95317; Eln.	
KW	Structural protein;	Repeat;	Signal; Connective tissue
FT	SIGNAL	1	27
			POTENTIAL.
FT	CHAIN	28	860
SO	SEQUENCE	860 AA;	71955 MW; 00CBESAALIEDD7F1 CRC64
			ELASTIN.

Query Match	27.98;	Score	976;	DB	1;	Length	860;
Best Local Similarity	39.38;	Pred. No.	3.4e-37;				
Matches	323;	Conservative	36;	Mismatches	249;	Indels	214;
				Gaps	40;		

OY	11	GSMASGKGLGSGGG-----AGAAAAAAAAAGAGCGGCGYGG	46
Db	78	GTFGAGCGGGLGAGCPGAGLCAFPAGTFPPGAGALVPBGAAAGMAAAATKAAKAGAGLGGYGG	137
OY	47	LGSGCTSGRGL-GCGGAGAAAAAAGAGCGGCGYGGLGSGCTSGRGLGCGGAGAA	105
Db	138	V--PGGVGCGVPGGVG-----GVPBGVGVGGVGP---GVGGIGIGLGGSTGA	184
OY	106	AAAAAAG-GAG--QGCGGLGSGCTSGRGLGGGGA-----GMAAAAAAA	149
Db	185	WVPVGGAGIAGGCKPGKVPVGVGLPGVPGGVLPGTARPPGVCLPVPPTGTGVKAKAPG	244
OY	150	AGGAGG--GYGGLGSG-----GTSGGPGYG---PGQCTSGRGLGCG-----	187
Db	245	GCGAFSGTIPGVGPPGCGQPPVPLGPTPKAKPLPGCTGLPTYNKTLRYVAAGAGKRGYPT	304
OY	188	--GAGAAAAAAGAGAGGCGYG--GLGSGCTSGR-----GGLGCGAGAAAAAA	235
Db	305	GTVGSGQAAAAAATKAAATYGAAGAGVLPVGCGGCIIPGAGAIPIGIGIAGAGTPMAAAAK	364
OY	236	AAAGCA-----GCGGCGLGSGCTSGRGLGSGGAG	266
Db	365	AAAAAAATYGAAGCLVPGGPCVRLPGAGIPEVGGIIPGVGGLPGVGGGIGGPTGVGPGAV	424
OY	267	AAAAAAGAAAGCA---GCGGCGY---GLGSGCTSGRGLGCGAGAAAAAAGAGAG	319
Db	425	SPAAAAATKAAATYGARCGVGIPTYGVGAGGPPYGVGAGAGLGGASPPAAAAAAKAK	484
OY	320	GCGYGLGSGCTSGCG---GYGPG---QOTSRRGLGCGAGAAAAA-----AG-GAG	368
Db	485	YGAGAGAGLGLLVGAVPALPGAVPVPAGVPGAGTPAATAAAAAAAAAAKAGLGPB	544
OY	369	QGCG-GGLGSGCTSGRGLGSGCGAGAAAAAAGAGCGGCGYGGLGSGCTSGRGLGSG	427
Db	545	VGCVPGCGVCGITGGVGVG-----VPGGVGPGVTTGIG---AGPGLGGGA	588
OY	428	GAGAAAAA-----AAAG-GAGCGYG--GLGSGCTSGRGLGCGAGAAAAAA	476
Db	589	GSPAAAKAAKAAKAAKAYRAAAGLGAAGVPFGAGAGVPGFAGAGVPGFAGAGVPGFGA	648
OY	477	AAG-----GAG-----GCGYGLGSG--GTSGPGCYPPCGQOTSRRGLGSGG--	516
Db	649	GAGVPGFAGAVPGSLAASKRAAKGAAGGGLGGRGLGPGGLG-----GPGGLGAGV	702
OY	517	---AGAA-AAAAAAGAGAGCGGCGYGLGSGCTSGRGLGSGGAGAAAAAAGAGAGCG	572
Db	703	GRVGAAPAPAAAAAAKAAKAAKAYGLGAGAGLGAAGLGAAGL-----AGGLGAG	753
OY	573	GY--GGLGSGCTSGRGLG--GQAGAAAAAAGAGAGCGGTGGLGSGCTSGRGLGCG	629
Db	754	GIGAGGLGAGCL-GAGGIGAGAGGVSPAAAAKAAATYGAAGLGA--LCARPPGCGVAARP	810
OY	630	GAGAAAAAAGAGAGCGG-----YAG-LGSGCTSGPGCYG	665
Db	811	GFGSLPTIPGGGAGLGVGKPPPYGAGLALGALGYCGGCGFCG	852

```

RESULT 10
ELS_RAT
ID      ELS_RAT      STANDARD;      PRT;      864 AA

```

DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 01-OCT-2000 (Rel. 40, last annotation update)
DE ELASTIN PRECURSOR (TROPOLASTIN) (FRAGMENT).

0S *Rattus norvegicus* (Rat).
 0C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 0C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91104868; Pubmed=1702999;

```



```

Oy 612 YGGLSGCTSGRGLG-----OGAGAAAAAAGG-----AG 646
Db 447 STPGGCGDGGRRKKGGFEGKHKGGSNPKFENIAGLRALLARSVERTTDECTWAG 506
Oy 647 QGGYGLG-----OGTS-----GPGY--GPGQGTSGIR 674
Db 507 VFVYGGSKTSLYNLRCTALAIPOCRLTPLSRPLFGMAPGPGPGLR 555

RESULT 15
ELS_CHICK STANDARD; PRT; 750 AA.
AC P07916;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN ELN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87242320; PubMed=3593675;
RX Bressan G.M., Argos P., Stanley K.R.;
RT "Repeating structure of chick tropoelastin revealed by complementary
RT DNA cloning.";
RL Biochemistry 26:1497-1503(1987).
RN [2]
RP SEQUENCE OF 85-750 FROM N.A.
RX MEDLINE=88309083; PubMed=2841924;
RX Baule V.J., Foster J.A.;
RT "Multiple chick tropoelastin mRNAs.";
RL Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
RN [3]
RP SEQUENCE OF 457-750 FROM N.A.
RX TISSUE=Aorta;
RX MEDLINE=87297534; PubMed=3502711;
RX Tokimitsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;
RT "Sequence analysis of elastin cDNA from chick aorta and
RT tissue-specific transcription of the elastin gene in developing chick
RT embryo.";
RL Arch. Biochem. Biophys. 256:455-461(1987).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NUCLEAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST TWO FORMS OF CHICKEN
CC ELASTIN THAT ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; M18633; AAA48761.1; -
CC EMBL; M21880; AAA49082.1; -
CC EMBL; M15889; AAA49108.1; -
CC PIR; A27264; A27264.
CC PIR; A26601; A26601.
CC HSP; P04002; IWER.
CC Structural protein; Connective tissue; Repeat; signal;
CC Alternative splicing.
CC NON_TER 1 24
CC SIGNAL <1 24
CC CHAIN 25 750 ELASTIN.

```

```

FT DOMAIN 83 686 8 X TANDEM REPEATS.
FT REPEAT 83 127 1.
FT REPEAT 219 262 2.
FT REPEAT 263 318 3.
FT REPEAT 319 393 4.
FT REPEAT 394 482 5.
FT REPEAT 483 554 6.
FT REPEAT 555 619 7.
FT REPEAT 620 686 8.
FT MOD_RES 63 66 8.
FT MOD_RES 66 66 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 111 111 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 115 115 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 156 156 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 159 159 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 198 198 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 200 200 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 235 235 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 252 252 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 256 256 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 297 297 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 301 301 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 354 354 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 357 357 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 427 427 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 431 431 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 513 513 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 517 517 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 520 520 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 586 586 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 590 590 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 593 593 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 655 655 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 658 658 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 719 719 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 721 721 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 743 743 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 748 748 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 750 750 OXIDATIVE DEAMINATION (POTENTIAL).
FT VARSPLIC 212 212 G -> GAGGCGQPGPGLGVPPIKAPLP (IN EMBRYONIC ISOFORM).
FT VARSPLIC 501 501 G -> GAGGCGQPGPGLGVPPIKAPLP (IN EMBRYONIC ISOFORM).
FT CONFLICT 536 536 A -> G (IN REF. 3).
FT CONFLICT 571 571 G -> A (IN REF. 3).
FT CONFLICT 610 610 P -> A (IN REF. 3).
FT CONFLICT 654 654 A -> R (IN REF. 3).
FT CONFLICT 667 667 P -> R (IN REF. 3).
SQ SEQUENCE 750 AA; 63697 MW; E57ECD60C6EE55F CRC64;

Query Match 20.2%; Score 704.5; DB 1; Length 750;
Best Local Similarity 36.7%; Pred. No. 2,4e-25;
Matches 284; Conservative 27; Mismatches 268; Indels 195; Gaps 43;

```


[illegible][illegible]


```

Db 1353 AAAAAAAAAAGAGGGGCGWGDGGYGSDSAAAAAAAAAAAAAAAAAGAGSGGCGYGGYSDS 1412
QY 582 T-----SGRGGLG-----QGACAAAAAAGAGGAGGCGGGL 615
Db 1413 AAAAAAAAAAGAGGAGGCGWGDGGYGSDSAAAAAAAAAAAAAAAAAGAGSGGCGYGGY 1472
QY 616 GS-----QTSRGGLGCGAGAAAAA-----AGAGGCGY- 650
Db 1473 GSDSAAAAAAAAAAAAAAAAAGAGGAGGCGYGSDSAAAAAAAAAAAAAAAAAGAGGAG- 1531
QY 651 ---GGLGS-----QTSRGCGYGPQGTGSTRPAA 678
Db 1532 WGDGGYGSISAAAAAAAAAAGSAGAGRGDGGYGSSSAAAAAAAAA 1579

RESULT 5
ID 053559 PRELIMINARY: PRT: 1489 AA.
AC 053559;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PG88-FAMILY PROTEIN.
RV3514 OR MTW023.21.
CN Mycobacterium tuberculosis.
OS Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAI1751.1; -.
DR TuberculList; RV3514; -.
DR InterPro: IPR000084; -.
DR InterPro: IPR002173; -.
DR Pfam: PF00934; pf.1.
DR ProDom: PD001223; -.
DR PROSITE: PS00583; PERK_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;
SO

Query Match 35 98; Score 1254.5; DB 2; Length 1489;
Best Local Similarity 39.58; Freq. NO. 1.2e-73;
Matches 336; Conservative 36; Mismatches 267; Indels 211; Gaps 36;

```

```

QY 217 SGRGGLGCG-----GAGAAA-----AAAAAAGA-----GCGYGG 248
Db 496 ACYGGAGCGGGGAGACGACADADQPCATGCTGTFAGAGAGAGAGSSAGGTNGSGGAG 555
QY 249 LGSQ-----TSRGGLGCGAGAAAAA-----AGAGGCGYGG-LGSQTSR 295
Db 556 TGGGCGAGAGAGAGADNPRTGIGTGDDGTGGAAGAGGAGAAAGTGTGTMGNAVG 615
QY 296 GGLGCG-----GAGAAAAA-----AAAGAGCG-----GGYGLGS 327
Db 616 CGAGCGGCGGAGAGACADADQPCATGCTGTFAGAGAGAGAGSSSAGTNSGAGTGR 675
QY 328 Q-GTSPGCGYGPQGTSGRGGLGCGAGAAAAA-----AGAGGCGYGG-LGSQTSRGCG 385
Db 676 QSGTGAGAGAGADNPRT-GTGTGCGDGTGTAAGAGAGAGAGAACTGTGCMITGTGNAVG 734
QY 386 LGGGAGAAAAA-----AGAGGCGYGG-LGSQGT-----SGRGGLGCGAGAAAAAAG 442
Db 735 AGSSGAGCGTNGSGAGCGTDGGGAGAGAGADNPRTGIGTGDDGTGGAAGAGAG 794
QY 443 AGCGYGG-LGSQTSRGGLGCG-----GAGAAAAA-----AAAGA----- 481
Db 795 AGTGTGCMITGTGNAVGAGAGCGGDDGAGAGADADQPCATGCTGTFAGAGAGAGSGG 854
QY 482 -----GCGYGLGSGQTSGPCC-YGPQQ 505
Db 855 SSCAGGTNGSGAGGTCGQVAVGAGISFNSGNSGCTGTGTGGV-GTGGDGNAGTGG 912
QY 506 TSGRGGLGCGAGAAAAA-----AGAGGCGYGG-LGSQTSRGGLGCGAGAAAAA 564
Db 913 DPKGCGTGTG-GTGSGGAGSGGAGNENGTCGTG-GTGKGLGTLNDGLSATVSTGG 969
QY 565 AAGCAGCGCGGGLGSGQTSRGGLGCGAGAAAAA-----GCGYGLGCG 618
Db 970 TGTGTGKGTGTGAGDSDSAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
QY 619 GTSRGGLGCG-----GAGAAAAA-----AAAGAGCGGCGGLG-----S 655
Db 1030 GDSG-SGLGCGGPFAGAGAGKAGAGSSAGCGTNGSGAGAGCGGAGAGISFNSG 1088
QY 656 QGTSRGCGYGG 665
Db 1089 GGTGTGTGGV 1098

RESULT 6
ID 053557 PRELIMINARY: PRT: 1079 AA.
AC 053557;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PG88-FAMILY PROTEIN (FRAGMENT).
CN RV3512 OR MTW023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).

```

DR EMBL: AL022022: CA11749.1: -
DR Tuberculin: Rv3512: -
DR InterPro: IPR002202: -
DR PROSITE: PS00318: HMG_COA_REDUCTASE.2; UNKNOWN.1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CDBC74B97D CRC64;

Query Match 35.2%; Score 1228.5; DB 2; Length 1079;
Best Local Similarity 40.0%; Pred. No. 4.2e-72;
Matches 318; Conservative 25; Mismatches 303; Indels 149; Gaps 29;

OY 14 ASGRGLGSGGCAAAAAAGAGCGGCGGCGTSGRGLGCGCAAAAAA 73
DB 38 AGCAGAGAGTG-GTGCAGAGTGTGGGCGGCGGCGGCGGCGGCGG 94
OY 74 AGC-----AGCGGCGGCG-----SOGTSGRGLGCG 99
DB 95 KCGNGGDAKACTGSAFAGTGTGGCGGCGGCGGCGGCGGCGGCGG 154
OY 100 GA-GAAAAAAGCA-----GCGGCGGCG-----SOGTSGRGLGCG 137
DB 155 GCGDGAAGAGTGTAGAGAGCGGCGGCGGCGGCGGCGGCGGCGG 214
OY 138 -----GAGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 215 PGANGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 274
OY 181 RCGLGCGCGCAAAAAAGAGAG-OGGCGGCGGCG-----GTSRGLGCG 225
DB 275 TCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 334
OY 226 G--AGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 272
DB 335 CGNGGAGAGTGTAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 392
OY 273 AAGAGAG-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 325
DB 393 VCGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 451
OY 326 GSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 382
DB 452 GGAAGAGAGTGTAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 510
OY 383 RCGLGCGGCG-----AGAAAAAAGAG-AGCGGCGGCGGCGGCG 429
DB 511 GCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 570
OY 430 GAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489
DB 571 AGCGT 626
OY 490 G-----SOGTSGRGLGCGGCGGCGGCGGCGGCGGCGGCGGCG 542
DB 627 GCGNRSGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 686
OY 543 GTSRGLGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
DB 687 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 746
OY 598 AAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 657
DB 747 GGT 804
OY 658 TSGPGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 672
DB 805 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 819

RESULT 7
ID 053775 PRELIMINARY: PRT: 1306 AA.
AC 053775;

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE GRS-FAMILY PROTEIN.
GN RV0578C OR MTW039.16C.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacterineae: Mycobacteriaceae: Mycobacterium.
OX NCBI_Taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL021942: CA11749.1: -
DR Tuberculin: Rv0578C: -
DR InterPro: IPR000084: -
DR InterPro: IPR000209: -
DR InterPro: IPR002173: -
DR Pfam: PF00934: PE.1.
DR Prodom: PD001233: -1.
DR PROSITE: PS00583: PFYB_KINASES.1; UNKNOWN.2.
DR PROSITE: PS00138: SUBPTILASE_SER; UNKNOWN.1.
DR SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match 34.3%; Score 1197; DB 2; Length 1306;
Best Local Similarity 38.1%; Pred. No. 5.2e-70;
Matches 325; Conservative 38; Mismatches 296; Indels 194; Gaps 35;

OY 3 GSHHHHHG-----SMAAGCGGCGGCGGCGGCGGCGGCGGCGG 53
DB 388 GGSAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 445
OY 54 GRCGLGCGGCAAAAAAGAG-AGCGGCGGCGGCGGCGGCGGCGGCG 101
DB 446 GRCGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 505
OY 102 GAAAAAAGAG-----AGCGGCGGCG-----GSGTSGRGLGCG 139
DB 506 MASAPAAAGSPGNGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 565
OY 140 -----AAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 186
DB 566 GCGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 625
OY 187 OGAGAAAAAAGAG-----AGCGGCGGCGGCG-----GTSRGLG 221
DB 626 RGADATGPGGTGASGGRGCGGCGGCGGCGGCGGCGGCGGCGGCG 685
OY 222 LGGCGGAGAAAAAAGAG-AGCGGCGGCGGCG-----GTSRGLG 264
DB 686 AGSGCGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 745
OY 265 AGAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 323
DB 746 GAGCGAARPGVAGSGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 803
OY 324 GLGSG--GTSRGLGCGG-----GCGTSGRGLGCGG----- 352
DB 804 GKGGRGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 863
OY 353 -AGAAAAAAGAG-----GCGGCGGCGGCGGCGGCGGCGGCGG 394

D	b		864	NCGNNGPBGAGSGGAGATFPGATSSSTHHVNGCNGCNGNDHALSNGAAGCNGCG	923
O	y		395	AAAAAAAAAGAGAGGGYGGGLGSQGTSGRGLGGOAGAAAAAAAGAGGCGYG- ---	450
D	b		924	NCSLTSGCAGAGHGCNGCNASRCMCGDCGTGGAGCNAAGICNGAGAGNCGDGTGSDNP	983
O	y		451	---LDSQTSRNGCLGCGCGAGAAAAAAAAGGAOQGYGGGLGSQGT-----SG	496
D	b		984	GATITSGGRGCGGGYGGCGGSVAIGDACOCRGGA--GGTGTYGLRTGATGATGTFDAG	1041
O	y		497	PFGYPGQQTSGRGLGCGAGAAAAAAA-----GGAQGQYGGGLGSQGTSG	546
D	b		1042	ADHGHCNGCTGVCGTCGACGGCGCAGCAKRLSPGNNGSGAGCDGAGAG--GTGG	1099
O	y		547	RGLLGCGGA-GAAAAAAAAGAGAGGYYGGLGSQGTSGRGLGCGAGAAAAAAAAG	605
D	b		1100	TGCGDGRRAGHTLFSSLAGTGTGTGNGGCTGTG--GTGAGAGAGGTGTLTATGTAAG	1157
O	y		606	GAGGCGYGGGLGSQGTSGRGLGCGAGAAAAAAAAGGAOQGYGGGLG-----SQGTSG	660
D	b		1158	RAGNGVGSGGSLGSAFPGPTGGWG-GAGGTSTVSAGDDGGRGGRGDPLDASSGNGG	1216
O	y		661	PGGY-GPGQQTSG 672	
D	b		1217	DGHGCDGFRTAG 1229	
 RESULT 8 ID O53552 PRELIMINARY: PRT: 1381 AA.					
AC	O53552:				
DT	01-JUN-1998	(TREMBLrel, 06, Created)			
DT	01-JUN-1998	(TREMBLrel, 06, Last sequence update)			
DT	01-OCT-2000	(TREMBLrel, 15, Last annotation update)			
DE	GRGS-FAMILY PROTEIN.				
GN	RV3507 OR MTW023.14.				
OS	Mycobacterium tuberculosis				
OC	Bacteria; Firmicutes; Actinobacteria; Actinomaderidae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-H37RV.				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Davlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutler S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,				
RA	Taylor K., Whitehead S., Barrett B.G.;				
RT	"Deciphering the biology of Mycobacterium tuberculosis from the				
RL	Nature 393:537-544(1998).				
DR	EMBL; AL022022; CAAL17744.1; -.				
DR	HSSP; P00778; 20DL.				
DR	Tuberculist; RV3507; -.				
DR	InterPro; IPR000084; -.				
DR	InterPro; IPR002173; -.				
DR	Pfam; PF00934; PE; 1.				
DR	ProDom; PD001223; -; 1.				
DR	PROSITE; PS00583; PKB_KINASES_1; UNKNOWN_2.				
DR	SEQUENCE 1381 AA; 110624 MW; CA0967BD07F6482 CRC64;				
 Query Match 33.9%, Score 1182.5; DB 2; Length 1381; Best Local Similarity 38.8%; Pred.No. 4.7e-69; Matches 321; Conservative 32; Mismatches 275; Indels 199; Gaps 37;					
Oy	16	GRGGLGGGAGAAAAAACG-----AGCGYGGGLGSQ-----GTSNGG-----LGQQ	61		

Db	163	GNAGAGGAGGSGGAAGNGNGMTCFACGTGIGTGTGAPGAMGCGTGGNGCATTGCG	222
Qy	62	GAGAAAAAAGAGAGAGCGTGC-----TGSQSTSGRGLGCGGAGAAAAAAGAG	116
Db	223	GLGAGAGCGGTGTGTGTGTGNGGALLTGAGGVGAGGAGCGGTGTGAGGAAGAGTG--G	280
Qy	117	QGGVGGTGSQSTSGRGLGCGGAGAAAAAAG-----AGCGGAGTGSQSTSP----	169
Db	281	NGGAGGLEPMNG--GDDGAGAGCGGCGAAGDAABAAGTGGKGGGGGCTGGAGACPVLF	338
Qy	170	-----GGYGPQQSTSGRGLGCGG-----AGAAAAAAGAGAGCG--Y	208
Db	339	GHGAGGAGCGGCGGTGCGAGGAGGDDTTLAAGTGTGEGGTGAAGAGCAARGALITSGLA	398
Qy	209	GGLGSGQSTSGRGLGCGGAGAAAAAAG-----GAGCGGTGGLGSO-----GT	254
Db	399	GGVAGAGGTGTGTGTGTGAGDAAAAYVFGANGDGFAGGAGGNGTGGAAVTTGVAADGCT	458
Qy	255	SGRGLGSGQ-----GAGAAAAAAG-----AGAGAGCGGTG	287
Db	459	GKGKGTGAGGAGCAGDAGSTGNPGKGGDGTGAGAGGAGAACTGNGGHAAGTGGDGGST	518
Qy	288	GSQSTSGRGLGCG-----GAGAAAAAAGAGAGCGGTGGLGSGQSTG	331
Db	519	GGNGNGNGTGCYNGADNTLNPDPGAGGEPGAGAGAGAGGAGGPGTGTGTGNG--GNG	576
Qy	332	GGPGYGPQQSTSGRGC-----LGGGAGAAAAAAGAGAGCGGTGGLGSGQSTG	381
Db	577	GNGNGGNGGNGGNGGNAANNSTNAPVYGGEG--GAGDGGGAGGAGAAAGTGA--GSQGTG	633
Qy	382	GRGLGCGGAGAAAAAAGAGAGCGGTGGLGSGQSTSGR-----GGTGGGAGAAAAA	437
Db	634	GVGGDGGAGAG-----GGGAGAGTGNNGNPFYDGEAGSSGAGGNGGVGAGAGAN	683
Qy	438	AAAGGA-----GQGGYGGT-----SQSTSGRGLGCGGAGAAAAAAG	479
Db	684	GGTGGSGGNGDGGAGGATGAGAGNGTGTGTBPAGGTGAKGGGGGCGGAGAGGAGNGAGG	743
Qy	480	G-----AGCGGTGGLGSO-----GTSGPGGTGPQQSTSGRGLGCGGAG--	517
Db	744	GQGGNAGCGGAGGAGGAGAAVTPGDBGKAPHPGADAGSGCGDGGK--CG--GSSGTTGSGCAP	800
Qy	518	-GAAAAAAGAGAGGCGGTGGLGSGQST-----GGRGLGCGGAGAAAAAAGAGAGG	572
Db	801	IGGAGGTGGGAGGAGGAGGAGTGTATTTPGNGGNAAGDGN---GNNAGAGGCGGG	857
Qy	573	GYGGLGSGQSTSGRGLGCGGAGAAAAAAGAGAG--GGYGGTGSQSTSGRGLGCGGAG	631
Db	858	DFGCGTNTSGAGSGGNGG--NAGTAGSGAGAGCTGTGTLGNGNG--GNGCNGCGDGCNA	914
Qy	632	-GAAAA-----AAAAAGAGAGCGGTGGLGSO--GTSGPGGTGPQQSTG	672
Db	915	HGTGAGFVFPATSLPTPNGAGAGGNGTGSNGCAPAG--APGPTTGG	960
RESULT	9		
053215	PRELIMINARY;	PRT; 1660 AA.	
AC	053215;		
DT	01-JUN-1998 (TREMblrel. 06, Created)		
DT	01-JUN-1998 (TREMblrel. 06, Last sequence update)		
DT	01-OCT-2000 (TREMblrel. 15, Last annotation update)		
DE	GRS-FAMILY.		
GN	RV2490C OR MTV008.46C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37Rv.		
NC	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,		

RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hohnsby T., Jagals K., Kiroh A., Mclean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021246; CAA16067.1; -
 DR Tuberculist: RV2490c; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR000228; -
 DR InterPro: IPR002173; -
 DR Pfam: PF00934; PE; 1.
 DR ProDom: PD001223; -; 1.
 DR PROSITE: PS00583; PKB_KINASES.1; UNKNOWN_1.
 DR PROSITE: PS01287; RTC_UNKNOWN_1.
 SO SEQUENCE 1660 AA; 133124 MW; 3A889CE12C0FA945 CRC64;

Query Match 33.6%; Score 1173; DB 2; Length 1660;
 Best Local Similarity 35.7%; Pred. No. 2,2e-68;
 Matches 331; Conservative 36; Mismatches 281; Indels 280; Gaps 37;

OY 11 GSMASGRGLGGGAGAAAAAAGGA-----GCGGCG----- 46
 DB 451 GPVSGNGNGKGNCAHAPVAGCHGNGGAGNGGLVGDGAGCHGCGDAAGAYADMTAI 510
 OY 47 -LGSQGT-----SGRGGLGCG-GAGAAAAAAGAGAGCGGCGGCGS----- 88
 DB 511 FLGSSGTPTGDEGCGAGAGAGAGAHAGDAGAGAGNGGAGAGGAGCAHFNALVSD 570
 OY 89 -----GTSRGGLGCGGAGAAAAAAGAGAGCGGCGG----- 123
 DB 571 GCGNGDAGAGRCGDCGAG-GAGGDAAPAGRAGSGCGVGDGAGAGAGAPNGSGCRDM 629
 OY 124 -----GSGGTSGRGGLGCGGA-----GAAAAAA 147
 DB 630 AFKDDGAGAGDGPAGCGAGAGAGATGATGATVHSGNGGKGNADATVAG 689
 OY 148 AAAGAGCGGTGGL-GSGGTSGPGYG-----PG---QQTSGRG 182
 DB 690 ANGAGGAGNGGLVGDGAGAGDGSAAANGANGVGEDAGDTLSGPGEGSEANGGQG 749
 OY 183 GLGCGAGAAAAAAGAGAGAGCGGCGGCGGCGGCGGAGAAAAAAGAG 242
 DB 750 GVGCGGAGAGAGDAGAGSALGSGNGRGRDAGAG-----GAGGAGAGAGAGSVSDG 805
 OY 243 QCGYCGGLSGGTSGRGGLGCGGAGAAAAAAGGA-GGCGYGLGSGGTSGRGGLGCG 301
 DB 806 PCGKCGAGAGAGAGAGS-GCGGKAGSADSAEAVGAGCGKGGDGVG--GVCGDGGPGGD 862
 OY 302 G-ACAAAAAAGAGAGCGGCGGGL-GSGGTSGPGYG-----PGQTSRGAG 347
 DB 863 GGAGGAGAGCGVSHGCGVGGDGLGAGGNGGDPGCHGSDGDCGDPG--AGLGG 920
 OY 348 LGG-OGAGAAAAAAGGA-----GAGCGCGYGLGSG-----GTSGR--GG 385
 DB 921 LGGSGNGCTRAASGVNDASDHGPGSGGNGNGNGAGASVAGAGAGNGDGNAGVGGG 980
 OY 386 LGGGAGAAAAAAGAGAGAG-----GCGYGLGSG--TSGRGGLGCG-GAGAAAA 435
 DB 981 AGNGGAG 1040
 OY 436 AAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486
 DB 1041 DGGAGGAG 1100
 OY 487 GG-----LGSQGTSGPGYG-----PGQTS 507
 DB 1101 GGAAAPSGTGVSHGTGCGVGDGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158

OY 508 GRGGLGCG-----GAGAAAAAAGAGAGCGGCGGCGG 543
 DB 1159 GGGGAGGSGNGCTSAANGVADSKHGPLTGGDGVGNGAKAAAGDGGDGDG--GNAG 1216
 OY 544 TSGRGGLGCGGAGAAAAAAGGA-GGGYGG-----LGSQGTSGRGGLGCGG----- 593
 DB 1217 LFGDAGAGDGDGTAAELAGDAGAGGKGNADAGDGDGCGDGAHGLGLT 1276
 OY 594 -----GAAAAAAG 644
 DB 1277 VAGNGAG 1336
 OY 645 AGCGYGLGSGGTSGPGYCGPGQTS 672
 DB 1337 NGGGGTGGCGGAG 1364

RESULT 10
 O9NHW2 PRELIMINARY; PRT; 1884 AA.
 AC O9NHW2;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila madagascariensis.
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 CC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
 OX NCBI_TaxID=115969;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20156766; PubMed=10688794;
 RA Hayashi C.Y., Lewis R.V.;
 RT "Molecular architecture and evolution of a modular spider silk protein
 RT gene."; 287:1477-1479(2000).
 RL EMBL; AF218623; AAF36091.1; -
 DR InterPro: IPR000087; -
 FT NON-TER 1884
 SO SEQUENCE 1884 AA; 148738 MW; 27B6F45339FD20A5 CRC64;

Query Match 33.2%; Score 1159.5; DB 5; Length 1884;
 Best Local Similarity 39.7%; Pred. No. 1.8e-67;
 Matches 318; Conservative 34; Mismatches 296; Indels 153; Gaps 35;

OY 11 GSMASGRGLGGGAGAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 69
 DB 564 GLGIGRSGPGGVPG-GSGPGSIGPGSGGCGGGLGPGGSGGPGGSGGCGG 622
 OY 70 AAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
 DB 623 GPGGCGGSGPGGAGAGGPGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 681
 OY 121 GLGSGGTSGRGGLGCGGAGAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 173
 DB 682 GPGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 741
 OY 174 PGQTSRGGLGCGGAG-AAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 229
 DB 742 PG-GAGPGCGPGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 798
 OY 220 AAAAAAAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 287
 DB 799 GAGPGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 858
 OY 288 GSG-----GTSRGGLGCGGAG-AAAAAAAGAGAGCGGCGG 321
 DB 859 GGTITIEDLITVDANGPITISEELTIGGAGAGGVPGSGPGGCGGCGGCGGCGGCGGCGG 918
 OY 322 Y--GGLGSGGTSGPGYGPGQTSRGGLGCGGAG-AAAAAAAGAGAGCGGCGGCGGCGG 378

```

Db 919 SGPBGVGS-GSGPGGVGPG--GSGPGVGS-GGFGPGGCGSPGPGVGP 975
Qy 379 GTSRGGLGCGC-----AGAAAAAAGAGCGGCGGCGGCGTSGRGL 424
Db 976 GSGSGGAGAGAGSVPGPGPYPGSGSGGPGGAGPYGPGAG-GPYGPGGPGYGPAGAG 1034
Qy 425 GGCGAGAAAA-AAAAAGAGCGGTGGLGSGCTSGRGLGCGAG-----AAAAA 477
Db 1035 GCEGPGGAGAGPYGPGGPGGAGPGGAG-----PGAGGPGGPGGAGGAGSGG 1090
Qy 478 AGCAAGCGGTGGLGSGCTSGPGGPGGCGTSGRGLGCGGAGAAAAA-----AA 528
Db 1091 PGAGAGPGGAG-----PGGPGGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 1146
Qy 529 GAGAGCGGTGGLGSGCTSGRGLGCGAG-----AAAAAAGAGCGGCGGCGG 575
Db 1147 GSGSGGPGGPGGPGGAGGAGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAG 1206
Qy 576 --GLGSGGTS--GRGLGCGGAG-AAAAAAGAGCGGCGGGL-GSGCTSGRGLGCG 629
Db 1207 GAGPGGVTGGLRGAGAGGAGRGAGRGAGRGAGRGAGRGAGRGAGRGAGRGAG 1266
Qy 630 GAG-----AAAAAAGAGCGGCGGCGG-----655
Db 1267 GSGGTTVIEDLDTIDADPTTISELTTISGAGGSGPGGAGPGGAGPGGAG 1326
Qy 656 OGTSPPGCGYGPQGTSGIRP 676
Db 1327 PGSGPGGAGPGG--GAGGPGYR 1346

RESULT 11
002402 PRELIMINARY: PRT: 738 AA.
ID 002402:
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE INSOLUBLE PROTEIN.
OS Pinctada fucata.
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pterioidea;
OC Pterioidea; Pteritidae; Pinctada.
OX NCBI_TaxID=50426;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97320490; PubMed=9177341;
RX Sudo S., Fujikawa T., Nagakura T., Ohkubo T., Sakaguchi K., Tanaka M.,
RA Nakashima K., Takahashi T.;
RT Structures of mollusc shell framework proteins."
RL Nature 387:563-564(1997).
DR EMBL; D86074; BAA20466.1;
SQ SEQUENCE 738 AA: 61723 MW: FDF984139BF3BA59 CRC64;

```

Query Match 33.1%; Score 1157; DB 5; Length 738;
 Best Local Similarity 43.8%; Pred. No. 1.2e-67;
 Matches 330; Conservative 30; Mismatches 191; Indels 202; Gaps 27;

```

Db 241 AAAAAAGAGGLGGLG-----GLGLGGLG--LGGLGGLGGLGGSMAAAAAA 294
Qy 204 GCGGTGGLGSGCTSGRGLGCGAG-----AAAAAAGAGCGGCGGCGGCGTSGRGC 259
Db 295 GGGGLGVGFY--GRGGRGRGRRRRAAAAAAAGAGGCGG-----GGCG 344
Qy 260 LGGCGAGAAAAAAGAGCGGCGG-----LGSCTSGR-----295
Db 345 GGGGAGAAAAAASASASRSGTDLGDIKDLRSNGASAKKASAVASTKS 404
Qy 296 -----GGLGCGAGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 446
Db 405 QIDDLKDVLDLGLLKSASASASASASAGCGGCGGCGGCGGCGGCGGCGGCGG 446
Qy 343 SGRGLGCGGAGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 402
Db 447 -GNGCGGCGGAGALAAALAA--GAG-----GGLGCGGCGGAGALAAALAA 488
Qy 403 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 462
Db 489 -GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 535
Qy 463 GCGAGAGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 522
Db 536 RGGGSAAAAAAAGAGGCGG--GCGG-----GFGVGLGCGGCGGCGGCGGCGG 585
Qy 523 AAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 581
Db 586 AAAAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 628
Qy 582 TSGRGLGCGGAGAAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 634
Db 629 -----GGSADVAAAAAAMYDGDADPFDMDGFCGNGNGGCGGCGGCGGCG 678
Qy 635 AAAAAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 667
Db 679 -GSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 710

RESULT 12
053395 PRELIMINARY: PRT: 1538 AA.
ID 053395:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PGRS-FAMILY PROTEIN.
GN PE_PGRS OR RV3345C OR MTY004.01C-MTY016.45C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RX Gordon S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Cole S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sultson J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RL Nature 393:537-544(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Parkhill J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021841; CAA17117.1;

```


DR HSSP; P00441: 1SOS.
 DR Tuberculist: RV3345C; -
 DR InterPro: IPR000084; -
 DR InterPro: IPR002173; -
 DR Pfam: PF00934; PE; 1.
 DR ProDom: PD001223; -; 1.
 DR PROSITE: PS00583: PKR_KINASES_1; UNKNOWN_3.
 DR SQUENCE 1538 AA; 129386 MW; 788F0B205587592 CRC64.

Query Match 33.1%; Score 1157; DB 2; Length 1538;
 Best Local Similarity 34.4%; Pred. No. 2, 2e-67;
 Matches 317; Conservative 36; Mismatches 308; Indels 260; Gaps 31;

QY 1 MRGSHHHHMSMASGNGGLGCGAGAAAAAAGG-----AGCGYGLGSGQTS 53
 DB 549 LNCAGAGAGTAPTSGGNGGAGATPTVAGENGAGAGNGHGSGVNGAGAGAGNVA 608
 QY 54 G-----RGGLGCGA-GAAAAAAGAGAGGCGYGLGSGQ-----GTS 91
 DB 609 GTGLALNGNGNGGICGNGGSAAGTGDDGKGNGAGAGAGQDFSAANGANGCGGNG 668
 QY 92 GRGGLGCGAGAAAAAAGAGAGCGCGYGLGSGQTSGRGLGCGA-----139
 DB 669 GNGIGCGGDAFATFAKAGNGAGNG-----GNVVAAGCGAGAGKGAIPAMKATGADG 724
 QY 140 -----GAAAAAAGAGAGCGCGYGLGSGQTSGRGLGCGA-----T 178
 DB 725 TAPTSGDNGNGNGASPTVAGCGGCGGSGNGNGAGAGAGAGAGAGAGAGAGPT 784
 QY 179 SGNGGLGCGAGAAAAAAGG--AGCGYGLGSGQTSGRGLGCGAGAAAAA 236
 DB 785 SGDSGTSGTGGAG 842
 QY 237 AAGGAGCGYGLGSGQTSGRGLGCGA-GAGAAAAAAGAGAGCGCGYGLGSGQTS 286
 DB 843 ANGANGENG-----GSGGNGGDDGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
 QY 287 -----LGSGQTSGRGLGCGA-----301
 DB 899 DGGAGENGLSGAMLPGCGTYVGNPTGCGNGGNGAGAGAGAGAGAGAGAGAGAG 958
 QY 302 -----GAGAAAAAAGG-----AGCGYGLGSGQTS 331
 DB 959 PNGNGNGNGNGAG 1018
 QY 332 GPGGY-----GPGQTSGRGLGCGAGAAAAAAGAGAGCGCGYGLGSGQTSGRGLG 387
 DB 1019 GKGNGNGAG 1078
 QY 388 GCGAGAAAAAAG 436
 DB 1079 GPGAGAGFASGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1138
 QY 437 AAAAG-----AGCGYGLGSGQTSGRGLGCGAGAAAAAAGAGAGAGAG 484
 DB 1139 NGGLAGNGGVSETFGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1196
 QY 485 GYGGLGSGQTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 532
 DB 1197 GKGNGNGD-----GNVGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
 QY 533 QCGYGLGSGQTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 588
 DB 1255 SCGAGNGGAG 1314
 QY 589 GCGAGAGAAAAA-----AAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 635
 DB 1315 GGNSSGAAG 1374
 QY 636 -----AAAAAAG-----AGCGYGLGSGQTS-----G 660
 DB 1375 DCGNGGNGNGNGSTVAGLAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434

QY 661 PGYGP-----QOTSGIRR 675
 DB 1435 NGAG 1455

RESULT 13
 Q9NHM4 PRELIMINARY; PRT: 2249 AA.

ID Q9NHM4; AC Q9NHM4; DT 01-OCT-2000 (Tremblrel. 15, Created) DT 01-OCT-2000 (Tremblrel. 15, Last sequence update) DT 01-Mar-2001 (Tremblrel. 16, Last annotation update) DE FLAGELLIFORM SILK PROTEIN (FRAGMENT). GN FLAG. OS Nephila clavipes (orb spider). OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae; Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila. OX NCBI_Taxid:6915; RN [1] RP SEQUENCE FROM N.A. RX MEDLINE=20156766; Pubmed=10688794; RA Hayashi C.Y., Lewis R.V.; RT "Molecular architecture and evolution of a modular spider silk protein gene."; RL Science 287:1477-1479(2000). DR EMBL; AF218621; AAF36090.1; -; DR InterPro: IPR000087; -; DR InterPro: IPR00209; -; DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1. FT NON_TER 1 FT TER 1 FT SEQUENCE 2249 AA; 174871 MW; 88C8B168A147CDA0 CRC64;

Query Match 33.0%; Score 1151; DB 5; Length 2249;
 Best Local Similarity 36.7%; Pred. No. 7, 5e-67;
 Matches 330; Conservative 30; Mismatches 288; Indels 250; Gaps 36;

QY 10 HGSMSGRNGGLGCGAG-----AAAAAAGAGAGAGAG-----GLGSGQTS 52
 DB 166 YPGGAGPGYGPAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG 225
 QY 53 -----SGRGLGCGAG-----AAAAAAGAGAGAGAG-----GLGSGTSGRG 94
 DB 226 VPGGAGPGYGPAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAG 285
 QY 95 -----GLGCGAGAAAAA-----AAAGAGAGAGAG-----GLGSGTSGRGAGAGAA 143
 DB 286 AGPGAG 345
 QY 144 AAAAAAG 170
 DB 346 AGRGAG 405
 QY 171 GYRGQTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 219
 DB 406 GAGPGAGVPGSRGPGAGAG-----GSGPGYVPGSRGPGAGAGAGAGAGAG 459
 QY 220 GGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265
 DB 460 GGVGTGAG 517
 QY 266 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
 DB 518 GAGGPGYGPAG 576
 QY 319 QGTY-----SGTSGPGYGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 356
 DB 577 PGYGPAG 636
 QY 357 -----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 396

```

Db 637 SCGYPCGSGPCGSGPCGSGPCGSGPCGSGPSTVPFGSGPBGSGPCGAGPCGAG 696
QY 397 AAAAAAGAGGCGY-GLSGSGTSGRGLGCGAG-AAAAAAAAAGAGGCGYGLGSG 454
   ||||| ||| : ||||| ||| : ||||| ||| :
Db 697 PCGAGGAGPCGAGVGLGAGRGAGRGAGSVAGRGAGRGAGRGAGRGAGRGAG 756
QY 455 GTSRGLGAGG-GAGAA-----AAAAAAAAAGAGGCG 485
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 757 GAGGAGGAGGPGGAGSGGTTVIEDLITIDADGPTTISELITISGAGSGPGAGTGG 816
QY 486 YGGLAS-OGTSPPGPGYPCQOTSGRGLGCGAGAA-----AAAAAAAAAGA- 531
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 817 VCPGSGPCGCVPCGFPBGCVPCGSGPGCVPCGAGRPRYRGCGSGPGAGAGTGA 876
QY 532 GCGG-YGGLGSGTSGRGLGCGAGAA-----AAAAAAAAAGA- 569
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 877 GPGGAGYPCGSGPGAGAGPGGEGPGAGGPGYGPBGAGPGYGPBGEGCGPYG 936
QY 570 -----GCGGYG-LGSQGTSGRGLGCGAGAAAAAAGAGCGGCGY-GLGSGG 619
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 937 PCGYSPGAGGPGYGPBGCGPGGAGGPGYGPBGYGPBGSGPGAGPGGYG 996
QY 620 TSGRG-----GLGSGGAGAAAAAAGAGGCGYGLGSG-OGTSPPGPGYPCQOTSG 672
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 997 PCGSGPCGTYGPGGSGPG-----GYGPGSGPGGTYGPGGSGPGGTYGPGGSGPG 1048

```

```

RESULT 14
O9GUB5 PRELIMINARY: PRT: 1468 AA.
ID O9GUB5:
AC 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DR 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HEAVY-CHAIN FIBROIN (FRAGMENT).
GN FIB-H.
OS Galleria mellonella (Wax moth).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Pyraloidea; Pyralidae; Galleriinae; Galleria.
OC NCBI_TaxID=7137;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=POSTERIOR SILK GLAND;
RA Zurevec M.; Kodrik D.; Yang C.; Sehna F.;
RT "Heavy-chain fibroin of Galleria mellonella L.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF095239; AAG10393.1; -.
FT NON_TER 1468 1468
SQ SEQUENCE 1468 AA; 122705 MW; 2DA59E1181BB3DDE CRC64;

```

Query Match 32.9%: Score 1148; DB 5; Length 1468;
 Best Local Similarity 30.1%; Pred. No. 8, 2e-67;
 Matches 354; Conservative 78; Mismatches 215; Indels 528; Gaps 36;

```

QY 14 ASGRGLGCGGGA-----GAAAAAAGAGGAGG----- 41
   ||| ||||| ||| : ||||| ||| : ||||| ||| :
Db 267 ASGLGLGAGAGSAGSAGAGLGVGVGSSSSAASAGAEVLLIDRSSASA 326
QY 42 -----GGYGLGSGTSGRGLGCGGAG-- 63
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 327 AAGSAGSVGGLGLGPGIGIPGIGATSTSGAGLGVGAGAGAGAG 386
QY 64 -----GAAAAAAGAGGAGG-----GGYGLG 86
   ||| ||||| ||| : ||||| ||| : ||||| ||| :
Db 387 ASGAGLGLGICAGSSSSAASAGAGAGVYIDRSSAASAGSAGLGLG 446
QY 87 SGTSGRGLGCGAGAGAA--AAAAAAGAGGCGYGLGSGTSGRG-----GLGSGGAG 140
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 447 LGLGLGPGYGLGVGVSASALGAGLGVGTAGASGLGAGAGVAVPAGAGLGVGAG 506
QY 141 AA-----AAAAAAGAGAGGCGYGLGSGTSGP-GGYGPG 175

```

```

Db 507 GSSGSSAASASARGPAPVYIEDGSSAASAAAAGSGAGLGLG-LGAWPPLGIGTN 565
QY 176 QOTS-----GRGGLGCGAGAGAA-----AAAAAGAG 204
   : : ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 566 EVSASATGSAAGTGAAGLGSAGSSAASAGAPVYIEDGSSAASAAAGSG 625
QY 205 QCGYGLGSGTSGRGLGCGG-----AGA 229
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 626 ASGLGLGLGAMGTLLGIGTPNGVSSASATGSAAGTIGSLCGSAGASLASAGAA 685
QY 230 A-----AAAAAAG-----GAGGCGYGLGSG 252
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 686 APVYIYIEDSSAASAAAGSGAGVGLGLGALPLGLGIPGICASSAGAGLGVGAA 745
QY 253 GTSRGLGCGGGA-----GAAAAAAGAGGAGG----- 281
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 746 GTSGLGIGTVASATGASAGAGLGLGAGSSGSSAASAGSAGAGVYIYIDRSSAAS 805
QY 282 -----GGYGLGSGTSGRGLGCGGAG 304
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 806 AAGSAGSAGPGLGLGVWGPLGIGIPGICASSASAGAGLGVGAAGTSGLGLGAGAS 865
QY 305 A-----AAAAAAGAGGCGG- 321
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 866 AAGSAGAGLGLGIGAGSSGSSVAASATGSGAGVYIYIDRSSAASAAAGSAGLGL 925
QY 322 -----YGGLGSGGTS-----GPGYGPQOTSGRGLGCGGAGAAAAAAGAG 367
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 926 GLGLGLPYGIGINGVSSAALGAGLGVG-----TAGAGLGLGLGTGASAGSAGAG 981
QY 368 GCGGYGLGSGTSGRGLGCGG-----AGAAAAAAGAGGAGGCG-----YG 412
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 982 GYGAGGSGFSSAASAGSAGAGVYIYIDRSSAASAAAAGSAGSGLGLGVWGPLG 1041
QY 413 LGSQGT-----GRGGLGCGGAGAA-----AAAAA 440
   : ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1042 IGPNGVSSATGSAAGTGAAGLGSAGSAGSAGSAGAPVYIYIEDGSSAASAAA 1101
QY 441 GGAGGCGYGLG-----SGTSGRGLGCGGAG----- 467
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1102 AGSAGAGLGLGLGAMGPLGIGTPNGVSSASATGSAAGSTG-AGLGSAGAGSAGS 1160
QY 468 GAA-----AAAAAAGAGGCGGYGLGSGTSGP-GGYG----- 502
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1161 GAAGPAPVYIYIEDGSSAASAAAGSAGSLGLG-LGAWGPLGIGTPNGVSSATGSA 1219
QY 503 GQOTSGRGLGCGGAGAA-----AAAAAAGAGGCGYGLGSGTSG 544
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1220 GSTGAGLGSAGSAGSAGSAGAPAPVYIYIEDGSSAASAGSAGVGLGLSL 1279
QY 545 SGRGLGCGGAGAGAA--AAAAAAGAGGCGYGLGSGTSGRG-----GLGCGAGAG-- 596
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1280 GPLGIGIPGIVSSASALGAGLGVGARFASGLGLGVNAGASAGAGLGVGAGCGSSG 1339
QY 597 -----AAAAAAGAGGCGGYGLGSGTSGRGLGCGGAGAA 634
   ||| ||||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1340 LSTASAGSAGAGVYIYINDRSSAASAAAGSAGLGLGLGIPGYGIGLNGVSSA 1399
QY 635 AA-----AAAAAAGAGGCGYGLGSGTSGPBG 667
   ||| ||| ||| : ||| ||| ||| : ||| ||| ||| :
Db 1400 SALGAGLGVGTAGAGLGLGTGTAGSAGAG 1434

```

```

RESULT 15
O06810 PRELIMINARY: PRT: 1329 AA.
ID O06810:
AC 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DR 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHEICAL 107.4 KDA PROTEIN.
GN PRS-FAMILY OR RV1450C OR MTCY493.04.

```

OS Mycobacterium tuberculosis.
 OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
 OC Actinomycetales: Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornby T., Jaffe K., Krogan A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL: Z95844; CAB09271.1; -.
 DR Tuberculist; Rv1450c; -.
 DR InterPro; IPR00084; -.
 DR InterPro; IPR002173; -.
 DR InterPro; IPR003536; -.
 DR Pfam; PF00934; PE: 1.
 DR PRINTS; PRO1370; TRNSINTIMNR.
 DR PRODOM; PD001223; -; 1.
 DR PROSITE; PS00583; PFRB_KINASES_1; UNKNOWN_2.
 DR Hypothetical protein.
 SQ SEQUENCE 1329 AA; 107410 MW; 777A125F6DBAB234 CRC64;

Query Match 32.8%; Score 1145; DB 2; Length 1329;
 Best Local Similarity 35.9%; Pred. No. 1.2e-66;
 Matches 323; Conservative 35; Mismatches 300; Indels 242; Gaps 35;

QY 11 GSMASGRGGLGCGGA-----GAAAAAAAAAGA-----GCGGCG 45
 DB 139 GGIWNGGAGSGGAPQVAGAGAGLFTGTGGAGAGAGAGSGGWLGNNGVG 198
 QY 46 GLSSGSGTSGRGLGCGAGAAAAAAGAGAGGCGGCGGCGGAGAGAA 105
 DB 199 GAGGQSLIG-GATGAGAGAGAGLFCVGTGPGCGPGCGVG--GTGAGGLGCTLYGAGG 255
 QY 106 AAAAAAGG-AGCGGCGGLGSG--GTSRGLGCGAGAAAAA-----AAAA 150
 DB 256 HGGAGGPGPIGVGHGAGGAGLGLVGGHGGAGHGAAGVAGAGEDLSPHGTSGVG 315
 QY 151 GAGAGGCGGLG--GCTSPGCGTSGRGLG-----GCGAGAAAAA 200
 DB 316 GDAAGDGTGGRGWLAAAGAGAGGAGGAGTGAAGFSAALIVAGDNGDPGAGAGGT 375
 QY 201 GGAGG--GGYGGLGSGTSGRGLGAGAGAAAAAAGG-----GCGGCGGLG 250
 DB 376 GGAAGTIGAHGAAGASPTSGAGAGAGAGAGHFSGGAAGAGAGAGGLVGNAGAGAG 435
 QY 251 SGGT-----SGRGLGCGAGAAAAAAGGAGAG--GCGGCGGLGSGTSGR 295
 DB 436 GNGAPGAPSPGDPNNGGAGAGAGAGGCGAGAGAGAGAGAGAGAGAGAGATGATGL 495
 QY 296 GGL-----GCGAGAAAAAAGG-----AGCA-----GCGG 322
 DB 496 NGLAGAGDGTGKGGNGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 555
 QY 323 GGLGSGT-----SGPGYGPQOOTSRRGLGAGG----- 353
 DB 556 GGNAGAKTTPNGCDGVGNGAGNGSRGIGAGGIGAGAGTAGADGARGATPTSGAGGCTG 615
 QY 354 --GAAAAAAAAAGAGAGCGGGL--GSGTSGRGLGCGAGAAAAAAGAGAG-- 407
 DB 616 GNGANATVAG 675
 QY 408 -GGYGGLGSG-----GTSRGLGCGAGAAAAAAGG-----AGCGG 448

DB 676 AGGAGAGAGRGCDRGDGGTSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 735
 QY 449 GGL-----GSGTSGRGLGCGAGAGAAAAAAGAGAGCGG--GLG 491
 DB 736 GGLAGNLFQNGTIGVGGSGGKGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 795
 QY 492 GGTSGPGGCGGCGQ-----TSGRGLGCGAGAGAAAAAAGAGAGCGGGL--GSG 542
 DB 796 GCGGAGAGTTPGAGAGAGTPTSGDGGDGGNGN-----SQVVGNGGDDGNGNGGSA 850
 QY 543 GTSRGLGCGAGAGAAAAA-----AAAGAGAGCGGCGGLGSGTSGRGG 587
 DB 851 GTGNGRGAGDGAAG 909
 QY 588 LGCGAGAGAAAAAAGAGAG--GGYGGLG-----SGTSGRGLGCG 629
 DB 910 LGFRGAG 969
 QY 630 G-----AGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
 DB 970 GAGPDAG 1026

Search completed: July 3, 2001, 15:05:26
 Job time: 591 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:08 ; Search time 86.32 seconds
(without alignments)
485.301 Million cell updates/sec

Title: US-09-490-291-6
Perfect score: 3543
Sequence: 1 MASMTGGCGMGRGSMASGRG.....TSGIRIRAPSTSPFHNNHHN 691

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_0601:*
2: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT:*
3: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT:*
4: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT:*
5: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT:*
6: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT:*
7: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT:*
8: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT:*
9: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT:*
10: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT:*
11: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT:*
12: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT:*
13: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT:*
14: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT:*
15: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT:*
16: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT:*
17: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT:*
18: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT:*
19: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT:*
20: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT:*
21: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT:*
22: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT:*
23: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	718	19	AAW53346
2	2138	60.3	718	21	AAV59070
3	2135	60.3	651	20	AAV40097
4	2129	60.1	718	12	AAK14308
5	2076	58.6	606	16	AAK99053
6	2076	58.6	606	20	AAV40100
7	2059	58.1	646	18	AAW27178
8	1972.5	55.7	606	16	AAK99055
9	1972.5	55.7	606	20	AAV40102
10	1971.5	55.6	604	16	AAK99057
11	1963.5	55.4	606	20	AAV40101

12	1405.5	39.7	831	16	AAK80168	pmSSI MSP spider
13	1400	39.5	615	20	AAV40099	Spider silk protei
14	1346.5	38.0	714	16	AAK99059	Spider dragline va
15	1346.5	38.0	714	20	AAV40103	Polymer of an anal
16	1241.5	35.0	1177	9	AAK80940	SLP111 protein com
17	1241.5	35.0	1177	11	AAK805307	SLP 111 (Silk-fibr
18	1241.5	35.0	1177	11	AAK95105	Silk 11k protein
19	1241.5	35.0	1177	18	AAW26342	Silk-11k protein
20	1241.5	35.0	1177	21	AAV78277	SLP111 amino acid
21	1241.5	35.0	1177	22	AAK72725	Repetitive protein
22	1241.5	35.0	1177	22	AAK63995	SLP111 protein seq
23	1241.5	35.0	1178	14	AAK41007	Silk-11k protein
24	1241.5	35.0	1178	14	AAW53518	Amino acid sequenc
25	1184.5	33.4	1023	19	AAW53524	Amino acid sequenc
26	1184.5	33.4	1059	14	AAK41013	SLP4 multimeric pr
27	1184.5	33.4	1059	18	AAW26348	SLP4 synthetic pro
28	1184.5	33.4	1101	21	AAV78283	SELP4 amino acid s
29	1180	33.3	1059	9	AAK82962	SELP4 protein comp
30	1175.5	33.2	980	16	AAK81318	Adhesion protein.
31	1157	32.7	738	19	AAW56163	New DNA sequence 1
32	1138.5	32.1	595	12	AAK14309	N.clavipes draglin
33	1138.5	32.1	595	19	AAW53347	Nephila clavipes s
34	1138.5	32.1	595	21	AAV59071	N. clavipes spider
35	1138.5	32.1	1038	17	AAK95107	Fibronectin cell b
36	1138.5	32.1	1038	22	AAK72727	Repetitive protein
37	1138.5	32.1	1038	22	AAK63997	FCB-SLP protein fr
38	1137.5	32.1	1332	17	AAK95109	Silk 11k protein
39	1137.5	32.1	1332	22	AAK72729	Repetitive protein
40	1137.5	32.0	1332	22	AAK63999	SLP-C protein sequ
41	1135	32.0	979	21	AAV78286	FCB-SLP111 amino a
42	1135	32.0	1050	18	AAW26350	Sequence encoding
43	1120	31.6	1066	11	AAK05312	Silk-11k protein
44	1116	31.5	1018	16	AAW01496	SLP-L2 polymer (sl
45	1108	31.3	946	11	AAK05309	

ALIGNMENTS

RESULT 1	
ID	AAW53346 standard; Protein: 718 AA.
AC	AAW53346:
DT	06-JUL-1998 (first entry)
XX	
DE	Nephila clavipes spider silk protein.
XX	
KW	Spider; Nephila clavipes; silk protein; tandem repeat; fibre; dragline;
KW	cocoon; tensile strength; elasticity.
XX	
OS	Nephila clavipes.
XX	
PN	US5728810-A.
PD	17-MAR-1998.
XX	
PF	19-APR-1995; 95US-0425069.
XX	
PR	15-APR-1991; 91US-0684819.
PR	20-APR-1990; 90US-0511792.
PR	04-OCT-1994; 94US-0317844.
PR	19-APR-1995; 95US-0425069.
PA	(UYWY-) UNIV WYOMING.
XX	
PI	Human MB, Lewis RV, Xu M;
XX	
DR	WPI; 1998-270437/24.
DR	N-PSDB; AAV23249.
XX	
PT	Recombinant spider silk proteins - useful for making fibres


```

QY 351 GCGGAGAAAA-----AAAAAGACGCGCGTSGTSGRGLGCG 392
DB 312 ggggagaaagagggg1gggagagagaaagagggg1gsgg -agrgg1gsgg 370
QY 393 AGAAAAAAGAGAGCGCGTSGTSGRGLGCGAGAAAAAAGAGAGCGCG 452
DB 371 ag---avaaaaagagggg1gsgg-----agrggagag-----aaaaagagagrgg 419
QY 453 LGSQGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 499
DB 420 lgnqg-agrgg1gsggagag-----aaaaagagagggg1gngagagrggaaaaagagag 475
QY 500 GGYGP-GQGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 552
DB 476 ggyg1gsggagag-----ggygagaaaaavagaggg1rggagagggg1gsgg--sgrg1 531
QY 553 GCGGAGAAAA-----AAAAAGAGAGCGCGTSGTSGRGLGCG 592
DB 532 ggggagaaagagaggg1gggagagagagaaagavrgg1gsgg-----agrgg 587
QY 593 GAGAGAGAGAGAGAGCGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGCGCG 652
DB 588 gggg-----aaaaagagagggg1gsggav-grgg1gsggag-----aaaaagagaggg 636
QY 653 GGLGSGTSGRGLGCGAGCGGTSGTSGRGLGCG 683
DB 637 ggygagagag-----aaagag1gsggag 659

```

RESULT 3
 AAY40097 standard; protein: 651 AA.
 AAY40097;
 19-NOV-1999 (first entry)
 Spider silk protein spiderolone major 1.
 Spider silk protein; spiderolone major 1; cosmetic; make-up;
 dermatological compositions; hair care; skin care; sunscreen;
 hormone; moisturizer; skin disorder; skin disorder.
 Nephila clavipes.
 FR2774588-A1.
 13-AUG-1999.
 11-FEB-1998; 98FR-0001614.
 11-FEB-1998; 98FR-0001614.
 (OREA) L'OREAL SA.
 Philippe M, Garson JC, Arraudau JP;
 WPI: 1999-510729/43.
 Cosmetic or dermatological composition containing spider silk protein,
 for hair or skin care, in make-up or sunscreens
 Claim 3; Fig 1; 32pp; French.

CC hormones, moisturizers or agents for treating disorders of the
 skin and hair.
 CC
 SQ Sequence 651 AA;
 Query Match 60.3%; Score 2135; DB 20; Length 651;
 Best Local Similarity 68.3%; Pred. No. 2, 1e-145;
 Matches 495; Conservative 15; Mismatches 39; Indels 176; Gaps 34;

```

QY 25 GCGGAGAAAAAAGAGAGCGCGTSGTSGRGLGCGAGCG 81
DB 1 gggg-----aaaaagagagggg1gsggagagggg1gsggagagagagagagagag 56
QY 82 GCGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 141
DB 57 ggyg1gsggagag-----agrggagag-----aaaaagagagggg1gsgg--agrgg1gsgg 107
QY 142 GAAAAAAGAGAGCGCGTSGTSGRGLGCGAGCG 201
DB 108 g-----aaaaagagagggg1gngg-----agrggag-----aaaa 142
QY 202 AGAGAGAGAGAGAGCGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGCGCG 252
DB 143 agagagggg1gsggag-agrgg1gsggagagagagagagggg1gsggagagggg1gsgg 201
QY 253 SGTSGRGLGCGAGAAAAA-----AAAAAGAGAGCGCGTSGTSGRGLGCG 291
DB 202 sqg-agrgg1gsggagagagagaggg1gggagagagagagagagagagggg1gsgg 260
QY 292 GCGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 350
DB 261 g-----agrggagag-----aaaaagagggg1gsggagagggg1gsggagrgg1 311
QY 351 GCGGAGAAAA-----AAAAAGAGAGCGCGTSGTSGRGLGCG 392
DB 312 ggggagaaagagggg1gggagagagagagagagggg1gsggagagggg1gsggagrgg1 370
QY 393 AGAAAAAAGAGAGCGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGCGCG 452
DB 371 ag---avaaaaagagggg1gsgg-----agrggagag-----aaaaagagagrgg 419
QY 453 LGSQGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 499
DB 420 lgnqg-agrgg1gsggag-----aaaaagagggg1gngagagrggaaaaagagag 475
QY 500 GGYGP-GQGTSGRGLGCGAGAAAAAAGAGAGCGCGTSGTSGRGLGCG 552
DB 476 ggyg1gsggagag-----ggygagaaaaavagaggg1rggagagggg1gsgg--sgrg1 531
QY 553 GCGGAGAAAA-----AAAAAGAGAGCGCGTSGTSGRGLGCG 592
DB 532 ggggagaaagagaggg1gggagagagagagagavrgg1gsgg-----agrgg 587
QY 593 GAGAGAGAGAGAGAGCGCGTSGTSGRGLGCGAGAAAAAAGAGAGAGCGCG 652
DB 588 gggg-----aaaaagagagggg1gsggav-grgg1gsggag-----aaaaagagaggg 636
QY 653 GGLGSGTSGRGLGCGAGCGGTSGTSGRGLGCG 683
DB 637 ggygag 641

```

RESULT 4
 AAR14308 standard; protein: 718 AA.
 AAR14308;
 15-JAN-1992 (first entry)
 N.clavipes dragline silk protein-1.
 DE


```
D 221 aggggylglsqg-agrgllyggagq-----aaagagagq-----gglgs 262
OY 303 QGAGAAAAAAGAGAGAGGCGTSGTSGPCGTCG-GOOTSGRGLGCGAGAAAAA 361
    ||||| | ||||| ||||| ||||| : ||||| | : || ||||| |
D 263 ggaaggaagaaagagagggylglsqg-aggggylglsqgag- -gggag- -a 314
OY 362 AAAAAAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGCGGCGTSGTSGT 421
    ||||| | ||||| ||||| ||||| : ||||| |
D 315 aaaaagagggylglsqg-agrgllyggagq-----aaagagagq----- 358
OY 422 GRCGCGGAGAAAAAAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAAAA 476
    ||||| | ||||| ||||| ||||| : ||||| |
D 359 -gglysgagaggaagaaagagggylglsqgagggylglsqgagagaa 416
OY 477 AAAAAAGAGCGGCGTSGTSGPCGTCG-GOOTSGRGLGCGAGAAAAAAGAG 528
    ||||| | ||||| ||||| ||||| : ||||| |
D 417 aaaaagagggylglsqgagggylglsqgagaaagagggylglsqgagagaa 476
OY 529 AAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGCGGCGTSGTSGTSG 583
    ||||| | ||||| ||||| ||||| : ||||| |
D 477 aagggggggylglsqgagggylglsqgagggylglsqgagaaagagggylglsqg 536
OY 584 TSGRGLGCGAGAAAAAAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAG 643
    : ||||| | ||||| ||||| ||||| : ||||| |
D 537 -agrgllyggagq-----aaagagagq-----ggllysgagagggagaa 578
OY 644 AGAGCGGCGTSGTSGPCGTCG 667
    ||||| | ||||| ||||| : ||||| |
D 579 agggagggylglsqg-aggggylg 601
```

RESULT 7
ID AAM27178 standard; Protein: 646 AA.

```
XX AC AAM27178;  
XX AC  
XX DT 09-DEC-1997 (first entry)  
XX DE Nephila clavipes spider silk protein.  
XX KW High strength film; fibre: woven article; parachutes; sails;  
XX KM Absorbent; body armour; heavy metal; biological weapon; chemical;  
XX KW Flavour; fragrance; Nephila clavipes.  
XX OS Nephila clavipes.  
XX PN WO9708315-A1.  
XX PD 06-MAR-1997.  
XX PF 22-AUG-1996; 96WO-US13767.  
XX PR 22-AUG-1995; 95US-0517694.  
XX PA (BASE/)/ BASEL R M.  
XX PA (ELIO/)/ ELION G R.  
XX PI Basel RM, Elion GR.  
XX DR WPI: 1997-179272/16.  
XX DR N-PSDB; AAT85356.  
XX PT New opt. multimerised DNA sequences encoding spider silk protein -  
XX PT making high strength films, fibres, woven articles etc.  
XX PS  
XX PS
```

Example 2; Fig 1; 57pp; English.

A process has been developed for the production of a DNA fragment encoding silk protein. The process involves: (a) selecting target DNA, from a silk-producing spider, that contains many repetitive and non-repetitive regions; (b) selecting a single-stranded DNA primer of at

least 10 nucleotides with a sequence that is complementary to a region of the target; (c) repetitively combining the primer with melted target DNA, incubating the mixture with nucleotides and a DNA polymerase with proofreading activity to produce a DNA fragment which is complementary to the target and is at least 2 kb long. The present sequence encodes the spider silk protein from Nephila clavipes. The DNA fragment can be used to make fibres, films, woven articles, e.g. for use in parachutes, sails, body armour, and absorbents (e.g. of heavy metals, biological weapons, DNA, chemicals, flavours and fragrances). The high molecular weight (90-250 kD) of spider silk proteins can be produced on a commercial scale (at over 2 g/l cell mass). It has better tensile strength and elasticity than silkworm silk. Inclusion of both repetitive and non-repetitive regions ensures isolation of stable clones.

Sequence 646 AA;

Query Match 58.1%; Score 2059; DB 18; Length 646;
Best Local Similarity 69.1%; Pred. No. 5,6e-140;
Matches 473; Conservative 18; Mismatches 58; Indels 136; Gaps 31;

```
OY 2 ASMTGQOMGRGSMASGRGLGCGAGAAAAAAGAGCGGCGTSGTSGRGLG 61
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
D 6 aaagagggggylglsqgagrgggag- -aaagagagggggylglsqg-agrglly 60
OY 62 GCGAGAAAAAAGAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGCGG 121
    ||||| | ||||| ||||| ||||| : ||||| |
D 61 gggag- -aaagagagagq-----ggllysgagaggaagaaagaggg 103
OY 122 YGGLGCGTSGRGLGCGAGAAAAAAGAGCGGCGTSGTSGPCGTCG-GOOT 180
    ||||| | ||||| ||||| ||||| : ||||| |
D 104 yggllysgq- -agrgggagq- -aaagagagggggylglsqg- -agggggylly 151
OY 181 SGRGLGCGAGAAAAAAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAG 240
    : ||||| | ||||| ||||| ||||| : ||||| |
D 152 agrgllyggagq- -aaagagagagq-----ggllysgagaggaagaa 194
OY 241 GAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGCGGCGTSGRGLG 300
    ||||| | ||||| ||||| ||||| : ||||| |
D 195 ggaagggggylglsqg-agrgggagagq- -aaagagagggggylglsqg-agrglly 248
OY 301 GCGAGAAAAAAGAGAGCGGCGTSGTSGPCGTCG-GOOTSGRGLGCGAGAAA 359
    ||||| | ||||| ||||| ||||| : ||||| |
D 249 gggagq- -aaagagagggagggylglsqg-agggggyllysgagagrgllyggag- - 301
OY 360 AAAAAAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGCGGCGTSGTSG 419
    ||||| | ||||| ||||| ||||| : ||||| |
D 302 aaaaagagagq-----ggllysgagaggaagaaagagggggyllyng 347
OY 420 TSGRGLGCGAGAAAAAAGAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAG 479
    | ||||| | ||||| ||||| ||||| : ||||| |
D 348 - -agrgggagq- -aaagagagggggyllysgq-agrgllyggagq- -aaaaa 394
OY 480 AGAGCGGCGTSGTSGPCGTCG-GOOTSGRGLGCGAGAAAAAAGAGAGCGG 538
    ||||| | ||||| ||||| ||||| : ||||| |
D 395 aggaagggggyllysgq-agggggyllysgagggyllyggagq- -aaaaaggggq- - 446
OY 539 GGLGCGTSGRGLGCGAGAAAAAAGAGAGCGGCGTSGTSGRGLGCGAGAAA 598
    ||||| | ||||| ||||| ||||| : ||||| |
D 447 - - - - -gglysgagaggaagaaagagggggyllysgq- - -agrgggagq- - 489
OY 599 AAAAAAAGAGAGCGGCGTSGTSGRGLGCGAGAAAAAAGAGAGCGGCGTSG 658
    ||||| | ||||| ||||| ||||| : ||||| |
D 490 - -aaagagagggggyllysgq-grgllyggagq- -aaagagagggggyllysg 540
OY 659 GTCGCGGTCGCGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSGTSG 683
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 541 asaa- - - - -saaasrlspgag 557
```

RESULT 8
ID AAR9055 standard; Protein: 606 AA.


```
Oy 404 AGCAGCGGCGGCGTSGRGLGCGAGAAAAAGAGCGGCGGCGTSGRG- 462
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 407 aaagp99y9p-9q9p99y9p-9q9p99y9p-9q9p99y9p-9q9p99y9p- 463
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 463 ---GLGCGGAG-----AAAAAAAAAGAGCGGCG-----GLGCGGTS----- 497
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 9q9p99y9p9q9p99p9saaaaa9p9q9p99y9p9q9p99y9p9q9p99y9p- 523
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 498 -----GPGGTCGCGGCGGCGGAGAAAAAGAGCGGCGGCGGCGTSGRG- 550
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 524 aaaaa9p99y9p9q9p99y9p9q9p99p9saaaaa9p9y9p-9q9p99y9p- 582
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 551 ---GLGCGGAG-----AAAAAAAAAGAGCGGCGGCGTSGRGLG-GCGAGAAAAA 602
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 583 9q9p99y9p9q9p99p9saaaaa9p9q9p99y9p-9q9p99y9p9q9p99p9sa 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 603 AAAAGCAGCGGCG-GLGCGGCGTSGRGLGCGGCGAGAAAAAG-----GAGCGGCGGCG 657
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 aaaaaa9p99y9p9q9p99y9p9q9p99p9saaaaa9p9y9p9q9p99y9p- 700
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 658 OCTSGGCGGCGGCG 671
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 701 p9q9p99y9p9q9 714
```

Search completed: July 3, 2001, 14:58:11
Job time: 441 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:34 ; Search time 48.42 Seconds

(Without alignments)
287.483 Million cell updates/sec

Title: US-09-490-291-6

Perfect score: 3543
Sequence: 1 MASMTGGCGMGKRGSMASRG.....TSGIRRAPSTFEHHHHH 691Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq Length: 0
Maximum DB seq Length: 200000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: /cgnt2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgnt2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgnt2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgnt2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgnt2_6/prodata/2/1aa/PCBUS_COMB.pep:*
6: /cgnt2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	718	1	US-08-425-069-2
2	2138	60.3	718	2	US-08-317-844B-2
3	2138	60.3	747	4	US-09-034-177-3
4	1405	39.7	832	1	US-08-209-747-2
5	1405	39.7	832	1	US-08-458-298-2
6	1241.5	35.0	1177	1	US-07-609-716-31
7	1241.5	35.0	1177	1	US-08-175-155-29
8	1241.5	35.0	1177	1	US-08-477-509B-64
9	1241.5	35.0	1177	2	US-08-707-237A-35
10	1241.5	35.0	1177	3	US-08-482-085B-64
11	1241.5	35.0	1177	4	US-08-475-411A-31
12	1241.5	35.0	1177	4	US-08-478-029A-31
13	1184.5	33.4	1059	2	US-08-175-155-48
14	1184.5	33.4	1059	2	US-08-707-237A-54
15	1184.5	33.4	1101	1	US-08-477-509B-83
16	1184.5	33.4	1101	3	US-08-482-085B-83
17	1157	32.7	738	3	US-08-864-038A-3
18	1138.5	32.1	595	2	US-08-425-069-4
19	1138.5	32.1	595	2	US-08-317-844B-4
20	1138.5	32.1	1038	1	US-07-609-716-36
21	1138.5	32.1	1038	4	US-08-475-411A-36
22	1138.5	32.1	1038	4	US-08-478-029A-36
23	1137.5	32.1	1332	4	US-07-609-716-41
24	1137.5	32.1	1332	4	US-08-475-411A-41
25	1137.5	32.1	1332	4	US-08-478-029A-41
26	1135	32.0	979	1	US-08-477-509B-89
27	1135	32.0	979	3	US-08-482-085B-89

28	1135	32.0	1050	1	US-08-175-155-54	Sequence 54, App1
29	1114	31.4	1018	1	US-08-089-862-11	Sequence 11, App1
30	1114	31.4	1018	1	US-08-587-333-18	Sequence 18, App1
31	1114	31.4	1018	5	PCT-US94-07776-16	Sequence 16, App1
32	1104	31.2	2100	3	US-08-477-509B-80	Sequence 80, App1
33	1104	31.2	2100	3	US-08-482-085B-80	Sequence 80, App1
34	1104	31.2	2107	1	US-08-175-155-45	Sequence 45, App1
35	1104	31.2	2107	2	US-08-707-237A-51	Sequence 51, App1
36	1103	31.1	907	2	US-09-010-928B-4	Sequence 4, App1
37	1099	31.0	945	1	US-08-089-862-6	Sequence 6, App1
38	1099	31.0	945	1	US-08-587-333-13	Sequence 13, App1
39	1099	31.0	945	5	PCT-US94-07776-11	Sequence 11, App1
40	1090.5	30.8	766	1	US-08-175-155-53	Sequence 53, App1
41	1090.5	30.8	766	1	US-08-477-509B-88	Sequence 88, App1
42	1090.5	30.8	766	2	US-08-707-237A-61	Sequence 61, App1
43	1090.5	30.8	766	3	US-08-482-085B-88	Sequence 88, App1
44	1084	30.6	745	2	US-09-010-928B-28	Sequence 28, App1
45	1081.5	30.5	870	2	US-09-010-928B-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-08-425-069-2
Sequence 2, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himan, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM: disk
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

Query Match 60.3%; Score 2138; DB 1; Length 718;
Best Local Similarity 66.3%; Pred. No. 3.5e-147;
Matches 498; Conservative 20; Mismatches 49; Indels 184; Gaps 35;
0Y 25 QGAGAAAAAAGAGAGCGTGGTGGT--SGRGIGGAG-AAAAAAAAAGAGAG 81

TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: N. clavipes
: TISSUE TYPE: minor amputate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
: US-08-209-747-2

Query Match 39.7%; Score 1405; DB 1; Length 832;
Best Local Similarity 50.1%; Pred. No. 2,1e-94;
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;

QY 23 GGCGAGAAAAAGAGAGCGGTGCGTSGRGLGCGGAGAAAAAGAGAGCG 82
DB 79 GGGYAG---AGAVAAAAAGAGAGY---RGAGYGGGGYGAAGAGAAAA--GAGAG 129
QY 83 GYGGLSGGTSGRGLGCGGAGAAAAAGAGAGCGY---GLGSGTSGRGLG 137
DB 130 GAGGY-----GRGAGAGAGAAAAAGAGAGAGYGGGGYGAAGAGAAAAAGAGAG 183
QY 138 GCGAGAAAAAGAGAGCGGTGCGTSGPGYGPQQTSRGLGCGGAGAAAA 197
DB 184 GYGRAGAGAGAAA--GAGAGYGGGGYGAAGAGAAAAAGAGSGAGYGRAGAGA 241
QY 198 AAAAAAGAGCGGTGCGTSGRGLGCGGAGAAAA-----AAAAG 241
DB 242 GAAAGAGAGAGY-----GGYGAAGAGAAAAAAXXXXXXXXXXXXXGAGAG 293
QY 242 GAGGAGY---GLGSGTSGRGLGCGGAGAAAAAGAGAGCGY--- 288
DB 294 YGGGAGAGAGAGAAAAAGAGAGAGYGRAGAGAGAAAAAGAGAGYGGGSGYAGA 353
QY 289 ---LGSGTSGRGLG---GCGAGAAAAAGAGAGCGGTGCGTSGPGYGP 341
DB 354 GAAAAAGAGAGAGYGRAGAGAGAGAGAGAGAGAGAGCGYGGGSGY---AGAGAGAA 412
QY 342 QGTSGRGLGCGGAGAAAAAGAGAGCGGTGCGTSGRGLG 390
DB 413 AAGAGAGAGYGRAGAGAGAGAGAGAGAGYGGGGYGAAGAGAAAAATGAGAGG 470
QY 391 OGAGAAAAAGAGAGCGGTGCGTSGRGLGCGGAGAAAAAGAGAGCGY 450
DB 471 YRGAGAGAGAGAGAGAGY---AGYGGGGYGAAGAGAGAAAAAGAGAGAGY 521
QY 451 G-GLSGGTSGRGLGCGGAGAAAAAGAGAGCGY---LGSGTSGPG-YGPG 505
DB 522 GRGAG---AGAGAAAAAGAGAGAGAGYGGGGYGAAGAGAGAGAGAGAG 577
QY 506 QGTSGRGLGCGGAGAAAAAGAGAGCGY---GLG---GTSRGLG-GGG 556
DB 578 YSRGAGAGAGAGAGAGAGAGAGYGGGGYGAAGAGAGAGAGAGAGAGAG 637
QY 557 AGAAAAAGAGAGAGAGCGGTGCGTSGRGLGCGGAGAAAAAGAGAGCGY 616
DB 638 AGAAAAAGAGAGAGAGY-----GGGGYGAAGAGAGAAAAAG--GAGRGY- 683
QY 617 LGSQGTSGRGLGCGGAGAAAAAGAGAGCGY 653
DB 684 -RGAGAGYGGGGYGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 717

,RESULT 5

US-08-458-298-2
: Sequence 2, Application US/08458298
: Patent No. 575677
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: TITLE OF INVENTION: CDNA Encoding Minor Amputate Spider
: TITLE OF INVENTION: Silk Proteins
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,298
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/209,747
: FILING DATE: 14-MAR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 832 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: NO
: FRAGMENT TYPE: Internal
: ORIGINAL SOURCE:
: ORGANISM: N. clavipes
: TISSUE TYPE: minor amputate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..309
: US-08-458-298-2

Query Match 39.7%; Score 1405; DB 1; Length 832;
Best Local Similarity 50.1%; Pred. No. 2,1e-94;
Matches 349; Conservative 17; Mismatches 207; Indels 124; Gaps 26;

QY 23 GGCGAGAAAAAGAGAGCGGTGCGTSGRGLGCGGAGAAAAAGAGAGCG 82
DB 79 GGGYAG---AGAVAAAAAGAGAGY---RGAGYGGGGYGAAGAGAAAA--GAGAG 129
QY 83 GYGGLSGGTSGRGLGCGGAGAAAAAGAGAGCGY---GLGSGTSGRGLG 137
DB 130 GAGGY-----GRGAGAGAGAAAAAGAGAGAGYGGGGYGAAGAGAAAAAGAGAG 183
QY 138 GCGAGAAAAAGAGAGCGGTGCGTSGPGYGPQQTSRGLGCGGAGAAAA 197
DB 184 GYGRAGAGAGAAA--GAGAGYGGGGYGAAGAGAAAAAGAGSGAGYGRAGAGA 241
QY 198 AAAAAAGAGCGGTGCGTSGRGLGCGGAGAAAA-----AAAAG 241
DB 242 GAAAGAGAGAGY-----GGYGAAGAGAGAAAAAAXXXXXXXXXXXXXGAGAG 293

ADDRESSEE: Flehr, Hohbach, Test, Abritton & Herbert
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: CA
 COUNTRY: US
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/175,155
 FILING DATE: 29-DEC-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Bertram I.
 REGISTRATION NUMBER: 20015
 REFERENCE/DOCKET NUMBER: A-55186-5/BR
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-781-1989
 TELEFAX: 415-398-3249
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1177 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-175-155-29

Query Match	35.0%	Score 1241.5	DB 1	Length 1177
Best Local Similarity	40.3%	Pred. No. 1.6e-82		
Matches 292	Conservative 58	Mismatches 313	Indels 61	Gaps 12

OY	7	GGQGRSMASRGELGGGCGAGAAAAAAGGAGCGGCTGCTGSGCTSGRGLGCGGAG	66
Dd	65	GSGGAGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG	124
OY	67	AAAAAAAAAAG-----GAGGGGCTGCGTSGRGLGGCGGAGAAAAA	113
Dd	125	SGAGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	184
OY	114	AGGAGCG--GTG-----GLGSGCTGGRGLGGCGGAGAAAAAAGAGCGGTGCTG	164
Dd	185	GAGAGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	244
OY	165	SGGTSGPGCTPFGGOT---SGRGLGGCGGAGAAAAAAGGAGCGGTGCTGSGTSGR	221
Dd	245	AGAGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	304
OY	222	G---GLGGGAGAAAAAAGAGAGCGGTGCTGSGCTSGRGLGGCGGAGAAAAA	278
Dd	305	GAGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	364
OY	279	GGAGCGGTG-----GLGSGCTGGRGLGGCGGAGAAAAAAGAGCGGTGCTGSG	331
Dd	365	AGSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	424
OY	332	TSGPGCTPFGGOT---SGRGLGGCGGAGAAAAAAGGAGCGGTGCTGSGTSGR--	386
Dd	425	GSGGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	484
OY	387	GLGGGAGAAAAAAGAGAGCGGTGCTGSGCTSGRGLGGCGGAGAAAAAAGAG	445
Dd	485	SGAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	544
OY	446	GCGGTG-----GLGSGCTSGRGLGGCGGAGAAAAAAGAGCGGTGCTGSG	498
Dd	545	GAAAGTGAAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS	604
OY	499	PGCTPFGGOT---SGRGLGGCGGAGAAAAAAGAGCGGTGCTGSGTSGAG---GL	555

[illegible]

```

1      RESULT      8
2      US-08-477-509B-64
3      : Sequence 64, Application US/08477509B
4      : Patent No. 5770697
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Ferrari, Franco A
9      : APPLICANT: Cappello, Joseph
10     : APPLICANT: Crissman, John W
11     : APPLICANT: Dorman, Mary A
12     : TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
13     : TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
14     : NUMBER OF SEQUENCES: 112
15     :
16     : CORRESPONDENCE ADDRESS:
17     : ADDRESSSEE: Flehr, Hohbach, Test, Albritton & Herbert
18     : STREET: Four Embarcadero Center, Suite 3400
19     : CITY: San Francisco
20     : STATE: California
21     :
22     : COUNTRY: US
23     :
24     : ZIP: 94111
25     :
26     : COMPUTER READABLE FORM:
27     :
28     : MEDIUM TYPE: Floppy disk
29     :
30     : COMPUTER: IBM PC compatible
31     :
32     : OPERATING SYSTEM: PC-DOS/MS-DOS
33     :
34     : SOFTWARE: Patentln Release #1.0, Version #1.30
35     :
36     : CURRENT APPLICATION DATA:
37     :
38     : APPLICATION NUMBER: US/08/477,509B
39     : FILING DATE: 07-JUN-1995
40     :
41     : CLASSIFICATION: 435
42     :
43     : PRIOR APPLICATION DATA:
44     :
45     : APPLICATION NUMBER: US 08/175,155
46     : FILING DATE: 29-DEC-1993
47     :
48     : PRIOR APPLICATION DATA:
49     :
50     : APPLICATION NUMBER: US 08/053,049
51     : FILING DATE: 22-APR-1993
52     :
53     : PRIOR APPLICATION DATA:
54     :
55     : APPLICATION NUMBER: US 07/114,618
56     : FILING DATE: 29-OCT-1987
57     :
58     : PRIOR APPLICATION DATA:
59     :
60     : APPLICATION NUMBER: US 06/927,258
61     : FILING DATE: 04-NOV-1986
62     :
63     : ATTORNEY/AGENT INFORMATION:
64     :
65     : NAME: Treccarlin, Richard F.
66     : REGISTRATION NUMBER: 31,801
67     : REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
68     :
69     : TELECOMMUNICATION INFORMATION:
70     :
71     : TELEPHONE: 415-781-1989
72     : TELEFAX: 415-398-3249
73     :
74     : INFORMATION FOR SEQ ID NO: 64:
75     :
76     : SEQUENCE CHARACTERISTICS:
77     :
78     : LENGTH: 1177 amino acids
79     : TYPE: amino acid
80     : STRANDEDNESS: single
81     : TOPOLOGY: linear
82     :
83     : MOLECULE TYPE: peptide
84     :
85     : US-08-477-509B-64

```


FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-029A-31

Query Match 35.0%; Score 1241.5; DB 4; Length 1177;
Best Local Similarity 40.3%; Pred. No. 1.6e-82;
Matches 292; Conservative 58; Mismatches 313; Indels 61; Gaps 12;

7 GQOMGRGSMASGRGLGSGGAGAAAAAGAGAGGCGGCGTSGRGLGCGGAG 66
65 GSGAGAGSAG 124
67 AAAAAAAAAA-----GAGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 113
125 SGAGAGSAG 184
114 AGGAGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 164
185 GAG 244
165 SGTGSGGCGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 221
245 AG 304
222 G---GLGSGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 278
305 GAGSAG 364
279 GAGAGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 331
365 AGSAG 424
332 TSGPGGCGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 386
435 GSGAG 484
387 -GLGSGGAG 445
485 SGAG 544
446 GCGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 498
545 GAAG 604
499 PGCGGCGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 552
605 AAG 664
553 GCGGAG 612
665 AG 724
613 GYG-----GLGSGTSGRGLGCGGAGAAAAAGAGAGAGAGAGAG 665

Db 725 GYGAG 784
QY 666 YGPG 669
Db 785 YGAG 788

RESULT 13
US-08-175-155-48
Sequence 48, Application US/08175155
Patent No. 5641648
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehr, Holbach, Teat, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175.155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-48

Query Match 33.4%; Score 1184.5; DB 1; Length 1059;
Best Local Similarity 40.7%; Pred. No. 1.8e-78;
Matches 272; Conservative 60; Mismatches 323; Indels 13; Gaps 7;

7 GQOMGRGSMASGRGLG-GGAGAGAAAAAGAGAGGCGGCGTSGRGLGCGG 65
384 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 443
66 GAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125
444 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 503
126 GSGTSGRGLGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 184
504 GSGAGAGSAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560
185 GLGSGAG 244
561 AG 620

[illegible]

Query Match	33.4%	Score 1184.5	DB 2	Length 1059
Best Local Similarity	40.7%	Pred. No. 1.8e-78		
Matches 272	Conservative 60	Mismatches 323	Indels 13	Gaps
<p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 07/114,618</p> <p>FILING DATE: 29-OCT-1987</p> <p>PRIOR APPLICATION DATA:</p> <p>APPLICATION NUMBER: US 06/927,258</p> <p>FILING DATE: 04-NOV-1986</p> <p>ATTORNEY/AGENT INFORMATION:</p> <p>NAME: Treacartin, Richard F.</p> <p>REGISTRATION NUMBER: 31,801</p> <p>REFERENCE/DOCKET NUMBER: A-55186-10/MHD</p> <p>TELECOMMUNICATION INFORMATION:</p> <p>TELEPHONE: (415) 781-1989</p> <p>TELEFAX: (415) 398-3249</p> <p>TELEX: 910 277299</p> <p>INFORMATION FOR SEQ ID NO: 54:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 1059 amino acids</p> <p>TYPE: amino acid</p> <p>STRANDEDNESS: unknown</p> <p>TOPOLOGY: unknown</p> <p>MOLECULE TYPE: protein</p> <p>US-08-707-237A-54</p>				
7	GOOMRGSMASRGGLG-GCGAGAAAAAAGAGAGGCGYGLGSGQTSGRGLGQGA	65		
384	GSAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG	443		
66	GAAG	125		
444	GSAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG	503		
126	GSQTSGRGLGQGAAG	184		
504	GSAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG	560		
185	GLGCGAG	244		
561	AG	620		
245	QCGYGLGSGQTSGRGLGQGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	304		
621	AG	680		
305	AG	363		
681	AG	737		
384	AAAG	423		
738	GSAG	797		
424	GLGCGAG	483		
798	GSAG	857		
484	GCGYGLGSGQTSGRGLGQGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	542		
858	GSAG	914		
543	SGQTSGRGLGQGAAG	602		
915	AG	974		
603	AAAAG	662		
975	AG	1034		
663	PGYGPQ 670			

This Page Blank (uspto)


```

QY 7 GQOMGRSMASGRGLGCG-----GAGAAAAAAGAGAGCGGGLGCG----- 52
Db 280 GGGGAGACAGSDGAGLGGTGTGTGAGAGAGCGRALLLLAGCGGGLGAGCGGTGCGAG 339
QY 53 -----GTSRGGGLGCG-----GAGAAA-----AAAAAAGC-AGCGGGLGCG 90
Db 340 GDEVLGAGVGTGCGKGGVAGAGAGAAAGQLFSAGAGAGAVGCTGGCGAGAGAA 399
QY 91 G-----TSGRGLGCGAGAAAAAAGAGAGCGGGLGCGSQTSGR-- 133
Db 400 GADAPASTGLTGTGFGAGAGGCGCGNAGAGGAGGAGCGGAGCGAGMCGSADN 459
QY 134 -----GGLGCGG-AGAAAAAAG-----AGCGGGLGCGSQT 168
Db 460 ASGITADGAGGTGCGNAGAGAGAACTGTGCVVAGAGAGAGCGGTGGAGAGAGT 519
QY 169 -----SGCGYGPQGTSGRGLGCGAGAAAAAAGAGAGCG----- 207
Db 520 DATATGATGTGTGFGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
QY 208 --GGYGLGCGSQTSGRGLGCGG--AGAAAAAAGAG--AGCGYGLGCGT--SGR 259
Db 579 GIGGTGCGTGGK--GAGAGAGCGGSSGAGGTNGSGAGGTGGCGAGAGAGAGADNPTGI 636
QY 260 GGLGCGAGAAAAAAGAGAGCGGTG-----LGSQGTSGRGLGCG--GAGAAAA 311
Db 637 GAGGTGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
QY 312 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355
Db 697 GSSATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
QY 356 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
Db 757 AGGAGGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
QY 394 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
Db 817 FDGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 436 -----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
Db 877 NGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
QY 476 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
Db 937 SINANGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 993
QY 532 GAGCGGTGGLGCGT-----GRCGLGCGAGAAAAAAGAGAGCGGGLG 580
Db 994 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
QY 581 SGTGSG--RGGLGCG--GAGAAA-----AAAAAAGAGAGAGAGAGAGAG 621
Db 1054 ATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
QY 622 --TSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
Db 1114 NFMGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169

```

```

RESULT 4
D70807
hypotheetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Colo, S.T.; Brosch, R.; Parhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Comor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.;
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

```

```

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:9825987
A:Accession: D70807
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; MIM:93261554; PDB:1CA17751.1; PDB:9292
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
C:Superfamily: collagen alpha 1(IV) chain

```

Query Match	35.68;	Score 1260.5;	DB 2;	Length 1489;
Best Local Similarity	38.58;	Pred. No. 4.3e-64;		
Matches 344;	Conservative 35;	Mismatches 274;	Indels 241;	Gaps 37;

```

QY 6 GQOMGRSMASGRGLGCG-----GAGAAAAA-----AA 35
Db 214 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 273
QY 36 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 90
Db 274 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 91 -----GTSRGGGLGCG-----GAGAAAAAAGAGAGCGGTGGLGCGSQT 131
Db 334 GTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 132 GRCGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175
Db 392 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451
QY 176 PGQGTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 227
Db 452 GADNPTGCGTGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
QY 228 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255
Db 512 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 571
QY 256 --TSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Db 572 NPTGIGTGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
QY 310 AAA-----AAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344
Db 632 ADADPGATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 691
QY 345 SGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
Db 692 -GIGGTGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 750
QY 404 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
Db 751 GGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 810
QY 460 GRCGLGCG-----GAGAAAAA-----AAAAAGAGAGAGAGAGAGAGAG 483
Db 811 GYAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 870
QY 484 -----GCGYGLGCGSQTSGRGLGCGAGAGAGAGAGAGAGAGAGAGAGAG 523
Db 871 CGQVYAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927
QY 524 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 582
Db 928 SGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 985
QY 583 GTSRGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 631
Db 986 SAGGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1044

```


Query Match	33.88%	Score 1196.5	DB 2	length 1381	
Best Local Similarity	38.7%	Pred. No. 1.6e-60			
Matches 319	Conservative 36	Mismatches 276	Indels 193	Gaps 36	
OY	7	GQMGRCGSMAS--	GRGGLGCGAGAGAAAAAAGC-----	AGCGCGTGLGSGT----	GTS 55
Db	150	GQSGGNGNGSAGLWNGGAGGAGGSGGAAGNGGNGWLFAGAGTGTGCTGATGAPGAMGSGTG			209
OY	56	GRGC-----	LGCGAGAGAAAAAAGAGAGCGGCGYG-----	LGSGGTSGRGLGAGGCGAGA	105
Db	210	GNGGNGALLIGGGGLGAGCGMGCTGGTGTGTGGGNGCAllLIGAGCGGAGGATGGGCGGA			269
OY	106	AAAAAAAAAGAGAGCGGCGYGLGSGGTSGRGLGCGAGAGAAAAAAGC-----	AGCGCGY		161
Db	270	GGAAGAGCGT--	GNGGAGGGLFPMNG--	GIDGAGGGGCGDCAAGDAAGTGGCGKGGGQGDG	325
OY	162	GLGSGGTSGRGGYPRGQGTSGRGLGCGAGAGAAAAAAGAGCGGCGYGLGSGGT---			218
Db	326	GTCGAGGAGPVLFFHG----	GAGMGCGG-----	GTCGAGGAGGAGGTTVI	366

Query March 33.3%; Score 1180; DB 2; length 1660;
Best Local Similarity 37.6%; Pred. No. 1.6e-59;
Matches 326; Conservative 38; Mismatches 222; Indels 212; Gaps 35;

OY 2 ASMTGGGQDMRGMSASGRGTGGCGAGAAAAAAGAC-----OGG 45
bb ::::|||||
514 SSGTGPGEDEGGNGG--AGGAGCAGAAHAGDGGAGAGNGNAGAGGNGAHGFNAVLYSDGG 572

QY 46 YGGLGSGTSGRGLGSGGAGAAAAAAGAGGCGTGL-GSQGTSGRGLGSGGAG 104
Db 573 NGGDD--GAGRGDGGAG-GAGGDAPAGRAAGSGGVDGAGGAGGAPGNGSGGSGGDM 629
QY 105 A-----AAAAAAGAGGCGGAGGGLGSGG-----TSGRGLGSGGAGAAAA 147
Db 630 AFKDDGAGAGDGDGDPGAGGAGGAGAGATGEGYTATATYHNSGNGSGKGNADATYAG 689
QY 148 AAAAAAGCA-----GQGGYGLGSGQTSRPGYGPQQ-----TSGRG 184
Db 690 ANGGKAGAGNGGLVGDGAGAGDGGGAAGANGANVGBDAGTLGSGPREGSEANGGQ 749
QY 185 GLGCGAGCAAAAAAAGAGAGCGGTGGLGSGQTSRGLGSGGAGAAAAAAGAGAG 244
Db 750 GVGGGGAGGAGAGDGGAGGSSALGSGGNGRGRDAGQAG---GAGGAGGAGGAGSVSGGG 805
QY 245 QGGYGLGSGTSGRGLGSGGAGGAGAAAAAAGAGCA-GGGYGLGSGQTSRGLGSGG 303
Db 806 PGGKAGAGAGAGAGAGAGAG--GGGKAGSGADSAEAVGAGAGAGGAGGAGGAGG 862
QY 304 G-AGAAAAAAGAGAGAGGAGGAGG-GSQGTSGRPGY-----PGQDTSGRG 349
Db 863 GGAGGAGAGAGYSHGTVGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 920
QY 350 LGG-AGAGAAAAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 387
Db 921 LGGDGNCTRAASGVADADHGGSGGNGNGNGAGAGAGAGAGAGAGAGAGAGAGGAGG 980
QY 388 LGGGAGAGAAAAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 437
Db 981 AGGAG 1040
QY 438 AAAAAAG 488
Db 1041 DCGAG 1100
QY 489 GC-----LGSQGTSGRPGY-----PGQDTSG 509
Db 1101 GGAAGTSGTVHGTGGVGGDGLGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
QY 510 GGGGAGGAGGAG 545
Db 1159 GGGGAGGAG 1216
QY 546 TSGRGLGSGGAG 595
Db 1217 LGGGAG 1276
QY 596 -----GAAAAAAG 646
Db 1277 VAGGAG 1336
QY 647 AGGGYGLGSGQTSRPGYGPQQTSG 674
Db 1337 NGGGTGGGAG 1364

RESULT 9
H70846
hypothetical glycine-rich protein RV3345C - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70846
R:Conor, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
N:ature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70846
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-1538 <COL>
A:Cross-references: GB:AI021841; GB:AI123456; NID:93261517; PIDN:COL17117.1; PID:9326
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV3345C
C:Superfamily: collagen alpha 1(IV) chain

Query Match 32.7%: Score 1158.5; DB 2; Length 1538;
 Best Local Similarity 34.7%: Pred. No. 2.4e-58;
 Matches 311; Conservative 34; Mismatches 305; Indels 247; Gaps 29;

```

QY      11 GKGSMASGKGGGLGGGGAGAGAAAAAAGG-----AGGGGGLGSGQTSG-----RG 58
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      557 GAAPTSGGNGGNGAGATPTTVAGENGAGCGNGCHGGSVNGAGAGAGNGVAGTGLALNG 616
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      59 GLGGGA-CAAAAAGAAAAAGAGAGGGGGLGSGQ-----GTSGRGLGSG 101
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      617 GAGNGNGIGGNGSAAAGCTGGDGKGGNGGAGANGPDFSASANGANGGGGNGGNGIGGK 676
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      102 GAGAAAAAAGAGAGAGGGGGLGSGQTSGRGGLGAGGA-----141
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      677 GGDAAATFPAKANGGAGGNG---GNVGAAGGGAGGKKAIPAMKATGADTAPTSGD 732
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      142 -----GAAAAAAGAGAGGGGAGG-LSGSGTSPGGGPGGQ-----TSGRGLG 188
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      733 GGGNGNGASPTYAAGNGNGDGGKGGSGGNGVNGNGNGAGNGAAGQACTGPTSGDSGTG 792
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      189 OGAGAAAAAAGAGAG-AGGGGGLGSGQTSGRGGLGAGAGAAAAAAGAGAGG 246
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      793 TDGAGAGNGAGAGAGAGTTLAGHCGNGKG--GNGGGGIGGAGERGADGAGPANGANGEN 850
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      247 GYGGLGSGQTSGRGLGSG-CAAGAAAAAAGAGAGGGGAG-----288
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      851 G-----GSGNGGDDGAGGAGGAKRAQAAGYTDGTTGTGGDGGNGDGGKAGDGGAGENG 906
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      289 -----LGSQTSGRGLGSGQ-----303
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      907 LMSGAMLPBGGTGYNPCTGCGNGNGSNAGVGTGKACTGSLTGLDGTGTPPANGNGN 966
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      304 ---GAGAAAAAAGAG-----AGGGGAG-----LGSQTSGPAGY---338
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      967 GGNCGKGTAGNGSGAAGNGNGSGLSGLNGDPAAGNGNGGALINQAFPGTGGKGGNGGN 1026
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      339 -GPGQQTSSRGGLGAGAGAAAAAAGAGAGGGGGLGSGQTSGRGLGCGAGAAA 397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1027 GAGAMTNGLGGFGGAGGGGAAVDVAATTGAGAGNGAGGFASTGTGLGPGGAGGPGAGDF 1086
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      398 AAAAAAAGAG-----GGGGYGGGLGSGQ-----TSGRGLGAGAGAAAAAAGAG--444
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1087 ASGVGAGAGAGDGGAGAGYGGFGGGGCTGCEBGRGTGANGSGSGDGGGGLSLGNGGLGNG 1146
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      445 -----AGGGYGGGLGSGQTSGRGLGAGAGAAAAAAGAGAGGGGGLGSG 494
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1147 GVSMTFGGAGGNGGYPGP--GPEGNGLGGNGGAGGNGGVSTTGGDGAAGKGGNGGD 1204
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      495 GTSPGCGYPPGQQTSSRGGLGAGAGAAAAAAG-----AGGGYGG 542
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1205 G--GNVGLGDDAGSGGAGNGGIGTDAGAGAGAGGAGGNGGSSKSTTTGMASSGGAGNG 1262
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      543 SGTSGRGLGAGAGAAAAAAG-----AAGAGAGAGGGYGGGLGSGQTS--GRGGLGAGAGAA 598
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1263 GTGLNAGAGAGAGAGAAAGYAFYFAGAAVGGDGGNGNGNGSHGGDGTTCGAGAGGAGNSSGA 1322
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      599 AAAA-----AAAGAGAGGGYGGGLGSGQTSGRGL-----GGGAGAAA-----637
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1323 AGSGGVYVNTAGHCGNGGNGNGNGNSAGAGGGGAGGAGNGGHGGAGTGDGCGNGNG 1382
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      638 -----AAAAAAG-----AGGGYGGGLGSGQTSGPAGYPPGQQTSG 674
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1383 GNSGNSTYAGLGLAGAAAGAGNGGGETSAAAGAGSGSGSGSTTTGAGAGAGNGAG 1439
  
```


Db 485 RGGAGCGSSGSGDGGGGAAGPAGW-----LTFDGDGAGNGGAAAAAGAGGQAGGGGGNGG 539
QY 369 -AGGGYGGTGGTSGT-----GRTGLGGGAGAAAAAAA---AAGGAGGGTGGTCSQ 418
Db 540 NGNGNGNGNGNGNGATGATGMLYNGAGAGGGGATAGAGCAGANGSVSTNGGTGNGMGIG-- 597
QY 419 GTSRGTGLGGGA--GAAAAAAAAGGAG--GGYGGTGG-----SGTSGRGTGGGQ--CAG 470
Db 598 GTGGSGGAGCAGAGLLGTGVCAGAGHGAGSGGAGDRTGGAGGTGFTSSDGAAGDGGDGGNGAG 657
QY 471 AAAAAAAAAAGCAGCGGTGGTGGTSGTSGPGYGPGRQRTSGRGTGLGGGAGAAAAAAA 530
Db 658 GTGGLTFGAGNGGGGGGGGAADTG--GNGAGNGCGGTGDNMGNGMSGG--AAGSGDG 712
QY 531 GGAGCGGCGTGLGSGGTSGRGTGLGGGAGAAAAAAAAGAGGGYGGTGGTSGRGTG 590
Db 713 GGAGGNGMFLFNGGAGGGGGGKGNAG-----GLGGSGFTGLPGLNGSGDGD 762
QY 591 GGCGAGAAAAAAAAGAGAGGGYGGTGGTSGRGTGLGGGAGAAAAAAAAGAGAGG 650
Db 763 GGNGAPGCVLLYNGGAG--GGGSSGGTGGPCAT--GGAGGKGGDGDQDLIGDGGNGNG 818
QY 651 GYGTGLSGGTSGPGGYPGPQOTSGL 675
Db 819 GAGGTG--CTPGPGPGGSGGLGTL 841

RESULT 12

A44112

spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)

N:Alternate names: silk fibroin, dragline

C:Species: Nephila clavipes

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998

C:Accession: A44112; S27824

R:Hirman, M.B.; Lewis, R.V.

J. Biol. Chem. 267, 19320-19324, 1992

A:Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes

A:Reference number: A44112; MUID:92406876

A:Accession: A44112

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-627 <HIN>

A:Cross-references: GB:M92913; NID:G159713; PID:G159714

A>Note: sequence extracted from NCBI backbone (NCBIF:113893)

R:Hirman, M.B.; Lewis, R.V.

submitted to the EMBL Data Library, May 1992

A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila clavipes

A:Reference number: S27824

A:Accession: S27824

A:Molecule type: mRNA

A:Residues: 19-627 <HIT>

A:Cross-references: EMBL:M92913

Query Match	32.18;	Score 1138.5;	DB 2;	Length 627;
Best Local Similarity	47.08;	Pred. No. 1.6e-57;		
Matches 325;	Conservative 36;	Mismatches 190;	Indels 141;	Gaps 36;

[illegible]

Db 159 SCPESSAAAAAAAAASGPQPGCGYGP - GQGGPGGYGPGQGPSPGSAAAAAAAAAASGPQ 217

Qy 243 AGCGGTGGLGSGGTSGRGGLG - GQ - - - - GAGAAAAAAAAAAGCAGCGGTGGLGSGGTSSGR 297

Db 218 QGPGGYG - - - - PGQGGPGGYGPGQGGTSGPSSAAAAAAGPGQGGPEGYGP - GQGGPSGP 272

Qy 298 GGLGGCGAGAAAAAAGAGAGCGGGYGGTSGGTSGPGGYPGQGTSGRGGLGCGGAGA 357

Db 273 G - - - - - SAAAAAAGAPGGYG - - - - PGQGGPGGYGPGQ - - - - - GPGSAGS 310

Qy 358 AAAAAAAAAAGCAGCGGTGGLGSGGTSGRG - - - - GLGGCGAGAAAAAAGAAAA - GAGCGGGY 412

Db 311 AAAAAAAAAAGPGQGGTGGYGP - GQGGPGGYGPGQGGPGGTGPGSSAAAAAAGPGQGGPGY 369

Qy 413 GGLGSGGTSGRGGLGCGGAGAAAAAAGAGAGCGGTGGLGSGGTSG - - RGLGCGGAG 470

Db 370 GP - GQGGSGSGG - - - - - SASAAAAAAGAPGGYGP - GQGGPGGYAPGQGGPSGP 417

Qy 471 AAAAAAAAAAGCAGCGGTGGLGSGGTSGPGGTGPGQGTSGRGGLGCGGAGAAAAAAGAAAA 530

Db 418 SASAAAAAAG - - - - - AAGPGYGP - - - - PGQGGPGGYAPGQ - - - - - GPGSPSSAAAAAAGAAAA 463

Qy 531 GAGAGCGGTGGLGSGGTSGRGGLGCGGAGAAAAAAGAGAGCGGTGGLGSGGTSGRGGL 590

Db 464 - - - - - AAGPGYGP - AQQGPGSPG - - - - - TAAASASGPGGYG - - - - - PAQGGP 500

Qy 591 GGCGGAGAAAAAAGAGAGCGGTGGLGSGGTSGRGGLGCGGAGAAAAAAG - - - - - AAGGA 647

Db 501 AGYCGGSVAASA - - - - - GAGSAGYGP - GSQSSAASRLASPDASARVASAVSNLVSST 555

Qy 648 GCGGTGGLGSGGTSGPGGYGPGQGTSGIRIA 679

Db 556 SSAALSSVTSNNAVSOIGASNPGLSGCDYLIA 587

RESULT 13
 A70812
 hypothetical glycine-rich protein Rv0833 - *Mycobacterium tuberculosis* (strain H37RV)
 C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70812
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentsles, S.; Hamlin, N.; Holtroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70812
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-749 <COL>
 A:Cross-references: GB:AL023004; GB:AL123456; NID:93261550; PIDN:CAAL17639.1; PID:9291
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv0833
 C:Superfamily: elastin

Query Match	31.58;	Score 1114.5;	DB 2;	Length 749;
Best Local Similarity	-39.68;	Pred. No. 4.1e-56;		
Matches 303;	Conservative 18;	Mismatches 309;	Indels 135;	Gaps 32;

[illegible]

A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAAL7640.1; PID:g291689
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: RV0834c
C:Superfamily: elastin

```
Query Match          30.6%; Score 1083; DB 2; Length 882;
Best Local Similarity 39.7%; Pred. No. 2,7e-54;
Matches 306; Conservative 28; Mismatches 306; Indels 130; Gaps 33;

QY 7 GQMGRCMSASGRGLGSGGAGAAAAAAGAGAGGCGYGLGSOGTSGRGGLSGGAG 66
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 133 GEAGCPGGMWLLGNGNGSGSAPGQTGAGAGAAAGLHGCTGAG--GTASGKGGTG 190
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 67 AAAAAAAAAAGAGAGCGYGLGSQ---GTSRGGLG--OGAGAAAAAAAAAGCA--G 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 191 LMSGGAGAGAGSG--GSGAGAGNLMFGTIGNGAGAGAAAGVNGVYAGAGAGALVA 249
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 119 QGGYGLGSQGTSGRGGLSGGAGAAAAAAGAGAGCGYGLGSQGTSGPGY--- 175
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 250 IGAGAGAGAAATTGTCGAGAGSN-ALGLFLGLSGGQGDSDMSAGSAGAGSGGAA 308
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 176 -PGQOTSGRGGLGCGAGAAAAAAGAG-----QGYYGLGSQGTSGRG 222
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 309 SPFGIDIGGAGGHG--GAGTNGAGAGAGAGSSGTFALDLSWAGAGNGAATTGTG 367
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 223 GLGCGAGAAAAAAGAGAGCGYGLG--SQTSRGGLSGGAGAAAA--AAA 275
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 368 GAG--CTGGFVAPDFIFGAAVYAGAGLGGATGAGTGTGCGYAGGFALGVYGA 425
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 276 AAGA-----GQGYGLGSQGTSGRGGLSGGAGAAAAAAG-----GAGQ 321
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 CGAGGAATETGIGGAGGLGVLLGAGAGGPG--GAASAGSGHGTGCDALGLIAGI 484
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 322 GGYGLG-----SQTSRPGCGYGPQOTSGRGLGSGGAGAAAAAAGAGCGYGG 376
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 485 GGVGVGGAAIDTGTGNGAGGSGTG-LLGVGAGAGHGG--ASVGTGSGGAGDGF 541
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 377 LGSQGTSGRGGLG-----GQAGAAAAAAGAGAGCGYGLGSQTSRGGLGQ 430
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 542 VAGAGNGAGTGYGVNANGNGSATGALAAVGA-----GAAGDATSGTGFG--G 594
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 431 AGAAAAAAGAGAGCGYGLGSQGTSGRGGLSGGAGAAAAAAGAG--QGYG 489
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 595 AGSARGLIFALGAGAA--GGDASTGVGPGPGGTASSPFCAIAIGAGAGCGAG 652
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 490 -----GLGSQGTSGPGYGPQ--QTSRGGLSGGAGAAAAAAG 528
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 653 TSGATGAGGDEVFEGIAVLGLFGAGAGGAATGDCATGAGGFGGAGAGIANFLGS 712
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 529 AAGAGCGYGLGSQGTSGRGGLSGGAGAAAAAAGAGAGCGYGLGSQTSGRG 588
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 713 VILHG--GAGAGGTAT--GTGNGAGAGGGGLSSPVILIGIGAGAGDGGALVIG--G 768
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 589 GLGCGAGAA-----AAAAAAGAGCGYGLGSQGTSGRG--GLGCGAGAAAAA 642
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 769 GDCGGGGEAVAVGIAVGAAGAGAAPTGNGAGGNGCDALIVGVGNGNAGTGTGAN 828
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 643 AAGGAGCG-----GYGLGSQGTSG--GPGY-----PGQ 670
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 829 TCGNGDPTIIVNMLAPSTLIGGNGNGVNGAGGTGKAGVFGAPG 878
```

Search completed: July 3, 2001, 14:59:40
Job time: 365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:39 ; Search time 37.35 Seconds

(Without alignments)

633.749 Million cell updates/sec

Title: US-09-490-291-6

Sequence: 1 MASMTCGQMGKSGMASGRC.....TSGIRIRAPSTSEHHHHH 691

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2138	60.3	747	1 SPDL_NEPCL	P19837 nephila cla
2	1281.5	36.2	1901	1 YZ08_MYCTU	O53553 mycobacteri
3	1174.5	33.1	5283	1 FBOH_BOMO	P05790 bombyx mori
4	1138.5	32.1	627	1 SPDL_NEPCL	P46804 nephila cla
5	1110.5	31.3	914	1 WA22_MYCTU	O06794 mycobacteri
6	1049.5	29.6	778	1 YQ34_MYCTU	P71933 mycobacteri
7	1032	29.1	957	1 Y278_MYCTU	P56877 mycobacteri
8	988.5	27.9	801	1 Y747_MYCTU	O53810 mycobacteri
9	976	27.5	860	1 ELS_MOUSE	P54320 mus musculu
10	939	26.5	864	1 ELS_RAT	O99372 rattus norv
11	869	24.5	603	1 YD25_MYCTU	O10637 mycobacteri
12	853.5	24.1	465	1 GRP2_PHAVU	P10496 phaseolus v
13	777.5	21.9	747	1 ELS_BOVIN	P04985 bos taurus
14	763	21.5	641	1 EBN1_EBV	P03211 epstein-bar
15	709	20.0	750	1 ELS_CHICK	P07916 gallus gall
16	691	19.5	730	1 ELS_HUMAN	P15502 homo sapien
17	686	19.4	498	1 Y118_MYCTU	O50615 mycobacteri
18	681	19.2	1733	1 VNUN_PPKVA	P33485 pseudocarie
19	678.5	19.2	543	1 YP91_MYCTU	O50603 mycobacteri
20	673	19.0	384	1 GRP1_PETHY	P09789 petunia hyb
21	669	18.9	481	1 LORI_MOUSE	P18185 mus musculu
22	656.5	18.5	1027	1 CAFE_RLEPA	P30754 riftia pach
23	644	18.2	338	1 GRP_ARATH	P27483 arabidopsis
24	630	17.8	1373	1 CA21_MOUSE	O01149 mus musculu
25	626	17.7	515	1 Y140_MYCTU	O50594 mycobacteri
26	622	17.6	1466	1 CA13_HUMAN	P02461 homo sapien
27	616	17.4	434	1 YK98_MYCTU	O10707 mycobacteri
28	608	17.2	1418	1 CA12_HUMAN	P02458 homo sapien
29	607	17.1	1464	1 CA13_MOUSE	P08121 mus musculu
30	606.5	17.1	1453	1 CA11_CHICK	P02457 gallus gall
31	603.5	17.0	1049	1 CA13_BOVIN	P04258 bos taurus
32	602.5	17.0	463	1 YA68_MYCTU	O53416 mycobacteri
33	602.5	17.0	1460	1 CA11_CANFA	O93937 canis famli

34	600	16.9	1366	1 CA21_HUMAN	P08123 homo sapien
35	599.5	16.9	1459	1 CA12_MOUSE	P28481 mus musculu
36	599.5	16.9	1464	1 CA11_HUMAN	P02452 homo sapien
37	599	16.9	1366	1 CA21_CANFA	O46392 canis famli
38	595	16.8	1364	1 CA21_BOVIN	P02465 bos taurus
39	592.5	16.7	1453	1 CA11_MOUSE	P11087 mus musculu
40	573.5	16.2	1362	1 CA21_CHICK	P02467 gallus gall
41	563	15.9	1758	1 CA24_CAUEL	P17140 caenorhabdi
42	561.5	15.8	1587	1 ICEN_XANCT	P18127 xanthomonas
43	559	15.8	1758	1 CA14_CAUEL	P17139 caenorhabdi
44	555.5	15.7	1779	1 CA11_BOVIN	P02453 bos taurus
45	553.5	15.6	1496	1 CA25_HUMAN	P05997 homo sapien

ALIGNMENTS

```

RESULT 1.
SPDL_NEPCL STANDARD; PRT: 747 AA.
ID SPDL_NEPCL
AC P19837;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).
OS Nephila clavipes (Ord spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_Taxid=6915;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 653-747 FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "Sequence conservation in the C-terminal region of spider silk
  proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and
  Araneus bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
CC - FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
  CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
  PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
  WITH ELASTIC AMORPHOUS SEGMENTS.
CC - SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.
CC - SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M37137; AAA29380.1; -
DR EMBL: U03848; AAB60212.1; -
DR PIR: A36068; A36068.
KW Sdk; Repeat.
FT DOMAIN 1 1 655
FT REPEAT 1 1 25
FT REPEAT 26 38
FT REPEAT 39 66
FT REPEAT 67 96
FT REPEAT 97 130
FT REPEAT 131 158
FT REPEAT 159 191
FT REPEAT 192 204
FT REPEAT 205 235
  
```

FT	REPEAT	236	262	10.
FT	REPEAT	263	292	11.
FT	REPEAT	293	305	12.
FT	REPEAT	306	333	13.
FT	REPEAT	334	360	14.
FT	REPEAT	361	394	15.
FT	REPEAT	395	424	16.
FT	REPEAT	425	458	17.
FT	REPEAT	459	485	18.
FT	REPEAT	486	512	19.
FT	REPEAT	513	525	20.
FT	REPEAT	526	535	21.
FT	REPEAT	556	582	22.
FT	REPEAT	583	612	23.
FT	REPEAT	613	642	24.
FT	REPEAT	643	655	25.
FT	CONFLICT	662	662	V -> L (IN REF. 1).
FT	CONFLICT	672	672	S -> T (IN REF. 1).
FT	CONFLICT	695	747	NGLSGCDVLIALLLEVSAIOLIGSSSIGOVNYSAGQAF FOIVGOSVYOAL -> ILVFLDVMSFRLFSRLFILLRS (IN REF. 1).
SO	SEQUENCE	747 AA:	60528 MW:	850E44B0BD49E012 CR664;

Query Match	60.3%	Score 2138	DB 1	Length 747
Best Local Similarity	66.3%	Pred. No. 3.2e-88		
Matches 498	Conservative 20	Mismatches 49	Indels 184	Gaps 35

[illegible]

Db 588 QGAG-----AAAAAGGAGAGGGYGGTGGGQGV-GRGGTGGGQAG-----AAAAGAGAGGGY 636

QY 653 GGTGSGGNGSGPFGYPCGQQTSGIRAPSTG 683

Db 637 GGVCSSGASAA-----SAAASRLSSPQAS 659

RESULT	2
YZ08_MYCTU	
ID	YZ08_MYCTU
STANDARD;	PRT; 1901 AA

DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sec

DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
DI 30-MAY-2000 (rel. 39, last annotation update)

05 *Mycobacterium tuberculosis*.
06 Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:

OC Actinomycetales;
OX NCBI_TaxID=1773;

SEQUENCE FROM N.A.
STPATN=H37PV.

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J.

RA Badcock K., Baslam D., Brown D., Chillingworth T., Connor R., Davies P., Foltwell E., Gontlee S., Hamlin N., Holroyd S.

RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA

RA Taylor K., Whitehead S., Barrell B.G.;
 re-isolated the histone of *Mycobacterium tuberculosis* from the

RT	complete genome sequence. "
RL	Nature 393:537-544 (1998).

CC -1- SIMILARITY: BELONGS TO THE MICROCOCCINAE SUBFAMILY, CC SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AL022022; CAA17745.1; -

DR Tuberculist; RV3508; -
DR Interpro: TPR000084: -

DR	Pfam; PF00934; PE; 1.
KW	Hypothetical protein; Repeat;

EM	CHAIN	31	1901	HYPOTHEETICAL PE-PGRS FAMILY PROTEIN
FT	31	1901		
DIAGRAM				

SQ SEQUENCE 1901 AA; 147627 MN; C7B1923D5D0146CD CRC64;

Query Match	36.2%	Score	1281.5	DB	1	Length	1901
Best Local Similarity	39.5%	Pred. No.	1.9e-50				
Matches	354	Conservative	28	Mismatches	273	Indels	241
						Gaps	39

QY 7 GQQMGRGSMASGRGLGQ-----GAGAAAAAAAAGAGAGCGGYGGLGSQ----- 52

Dd

280 GGGGGAGGAGSDGALDGTGGTGGTAGAAGAAGKALLLGAAGCGCCAAAGCCTGCATCATCA 333

[illegible]

91 G-----TSGRGGLGGGAGAAAAAAAGAGAGCGGTGGLGSGTSGR-- 133

— — — — —

```

Db 400 GADAPASTGLTGTGTFAGAGAGVGGGGMALAGINGSAGGTGGGAGMGSGADN 459
Qy 134 -----GGTGGG--AGAAAAAAGG-----AGGGGGLSGSGT 168
Db 460 ASGIGADGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 519
Qy 169 -----SGPGGYPGPGTGTSGRGLGAGAGAAAAAAGAGAGG----- 207
Db 520 DATATGATGCTGTGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 578
Qy 208 --GGYGLGSGQTSGRGLGAGG--AGAAAAAAGAG--AGGGYGLSGSGT---SGR 259
Db 579 GIGCTGTGTGGK--GAGGAGAGGSSAGGTGNSGAGGTGGGAGAGAGAGAGAGT 636
Qy 260 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 311
Db 637 GAGAGGTGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 696
Qy 312 AAAAAAGG--GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 355
Db 697 GSSATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
Qy 356 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 393
Db 757 AGGAGGTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
Qy 394 -----GAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 435
Db 817 FDGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
Qy 436 -----AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 475
Db 877 NGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
Qy 476 AAAAAAGG--GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
Db 937 SINANGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 993
Qy 532 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 580
Db 994 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1053
Qy 581 SGGTSG--AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 621
Db 1054 ATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
Qy 622 --TSGRGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 667
Db 1114 NPNCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1169

```

```

RP SEQUENCE OF 1-168 FROM N.A.
RX MEDLINE=80045039; PubMed=498286;
RA Tsujimoto Y., Suzuki Y.;
RT "The DNA sequence of Bombyx mori fibroin gene including the 5'
RT flanking, mRNA coding, entire intervening and fibroin protein coding
RT regions.";
RL Cell 18:591-600(1979).
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=79211211; PubMed=455439;
RA Tsujimoto Y., Suzuki Y.;
RT "Structural analysis of the fibroin gene at the 5' end and its
RT surrounding regions.";
RL Cell 16:425-436(1979).
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RX STRAIN=KINSHU X SHOWA;
RA Mita K., Ichimura S., Zama M., James T.C.;
RT "Specific codon usage pattern and its implications on the secondary
RT structure of silk fibroin mRNA.";
RL J. Mol. Biol. 203:917-925(1988).
RN [5]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94365842; PubMed=7916056;
RA Mita K., Ichimura S., James T.C.;
RT "Highly repetitive structure and its organization of the silk fibroin
RT gene.";
RL J. Mol. Evol. 38:583-592(1994).
RN [6]
RP SEQUENCE OF 5179-5263 FROM N.A., AND DISULFIDE BONDS.
RX STRAIN=J-139;
RA Tanaka K., Kajiyama N., Ishikura K., Waga S., Kikuchi A., Ohtomo K.,
RA Takagi T., Mizuno S.;
RT "Determination of the site of disulfide linkage between heavy and
RT light chains of silk fibroin produced by Bombyx mori.";
RL Biochim. Biophys. Acta 1432:92-103(1999).
CC - FUNCTION: FORMS THE SILK FILAMENT; A STRONG, INEXTENSIBLE,
CC INSOLUBLE AND CHEMICALLY INERT FIBRE.
CC - SUBUNIT: FORMED OF TWO CHAINS: HEAVY AND LIGHT, THAT ARE LINKED BY
CC A DISULFIDE BOND. HEAVY-LIGHT CHAIN ASSEMBLY IS ESSENTIAL FOR THE
CC EFFICIENT INTRACELLULAR TRANSPORT AND SECRETION OF FIBROIN.
CC - TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE POSTERIOR (PSG)
CC SECTION OF SILK GLANDS.
CC - DOMAIN: COMPOSED OF ANTIPARALLEL BETA SHEETS. THE STRANDS OF THE
CC BETA SHEETS RUN PARALLEL TO THE FIBER AXIS. LONG STRETCHES OF SILK
CC FIBROIN ARE COMPOSED OF MICROCRYSTALLINE ARRAYS OF (-GLY-SER-GLY-
CC ALA-GLY-ALA-)N INTERRUPTED BY REGIONS CONTAINING BULKIER RESIDUES.
CC THE FIBER IS COMPOSED OF MICROCRYSTALLINE ARRAYS ALTERNATING WITH
CC AMORPHOUS REGIONS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: AF226688; AAF76983.1; -
CC EMBL: V00094; CAA23432.1; -
CC EMBL: V00097; CAA23433.1; -
CC EMBL: S74439; AAB31861.1; -
CC EMBL: X13869; CAA32076.1; -
CC EMBL: M35378; AAA27839.1; -
CC EMBL: AB017362; BAA33147.1; -
CC PIR: S01844; S01844.
CC DR Silk; Signal; Repeat.
CC KW SIGNAL 1 21
CC FT CHAIN 22 5263
CC FT DOMAIN 149 5206
CC FT DISULFID 5244 5244
CC
CC POTENTIAL.
CC FIBROIN HEAVY CHAIN.
CC HIGHLY REPETITIVE.
CC INTERCHAIN (WITH LIGHT CHAIN).

```

FT DISULFID 5260 5263 C -> V (IN REF. 2).
 FT CONFLICT 10 10
 SQ SEQUENCE 5263 AA; 391586 MW; 8EE11D3A0A47440E CRC64;

Query Match 33.1%; Score 1174.5; DB 1; Length 5263;
 Best Local Similarity 39.6%; Pred. No. 1.7e-45;
 Matches 318; Conservative 49; Mismatches 305; Indels 131; Gaps 27;

2 ASMTGQGMGRS-MASGRGLGGGAGAAAAAAGAGAGGAGGGLGSGGTSGRGL 60
 DB 4337 AGAGAGTGAAGSAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4396
 QY 61 GGAGGAAAAAAGAGAGAGG---GAG---GLGSGGTSGRGLGGGAGAAAAA 112
 DB 4397 SCAG 4456
 QY 113 AAGGAGGAGGGLGSGGTSGRGLGGGAGAAAAAAGAGAGG---GAG---GLGSGGTSG 170
 DB 4457 SCAG 4516
 QY 171 PG---GYGPGQGTSG---RGGLG-----GGAGAGAAAAAAGAGAGAG 207
 DB 4517 AGAGAGAGAGTSSGFRPVYANGYSGYRYMSSESDFCTGAGAGAGAGAGAG 4576
 QY 208 G-----GAG-----GLGSGGTSGRGLGGGAGAAAAAAGAGAGGAGG 253
 DB 4577 GAG 4636
 QY 254 QGTSGRGLGGGAGAGAAAAAAGAGAGG-----GAG----- 287
 DB 4637 GAG 4696
 QY 288 -----GLGSGGTSGRGLGGGAGAAAAAAGAGAGGAGGGLGSGGTSGPG---GYGP 340
 DB 4697 GSGSAG 4756
 QY 341 GQGTSGRGLG-----GGAGAGAAAAAAGAGAGGAGGGLGSGGTSGRGLG 391
 DB 4757 GYGTGAG 4816
 QY 392 GAGAGAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGAGAGAGAGAG 443
 DB 4817 GAG 4876
 QY 444 GAGAGGAG 491
 DB 4877 GAG 4936
 QY 492 GSGGTSGPG---GYGPGQGTSGRGLGGGAGAAAAAAGAGAGGAGGGLGSGGTSG 548
 DB 4937 GAGYAG 4995
 QY 549 RGLGGGAG 606
 DB 4996 AAG 5055
 QY 607 GAGAGG-----GAG---GLGSGGTSGRGLGGGAGAAAAAAGAGAGG---GAG 653
 DB 5056 SGAG 5115
 QY 654 -GLGSGGTSGPG---GYGPGQGTSG 674
 DB 5116 AG 5138

RESULT 4
 SPID2_NEPCL STANDARD; PRT: 627 AA.
 AC P46804;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 PT 01-OCT-1996 (rel. 34, Last annotation update)

DE SPIDROIN 2 (DRAGLINE SILK FIBROIN 2) (FRAGMENT).
 OS Nephila clavipes (Ord spider).
 CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 CC Araneomorphae; Entelegynae; Araneioidea; Tetragnathidae; Nephila.
 OX NCI_TaxID=6915;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92406876; PubMed=1527052;
 RA Himan M.B., Lewis R.V.;
 RT "Isolation of a clone encoding a second dragline silk fibroin.
 RT Nephila clavipes dragline silk is a two-protein fiber.";
 RL J. Biol. Chem. 267:19320-19324 (1992).
 CC -1- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE
 CC CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF
 CC PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED
 CC WITH ELASTIC AMORPHOUS SEGMENTS.
 CC -1- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 1, OF THE DRAGLINE SILK.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M92913; AAA29381.1; -.
 DR HSSP; P04002; IWFA.
 KW Silk; Repeat.
 FT NON_TER 1 1
 FT DOMAIN 1 530 15 APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1 36
 FT REPEAT 37 79
 FT REPEAT 80 121
 FT REPEAT 122 172
 FT REPEAT 173 213
 FT REPEAT 214 252
 FT REPEAT 253 283
 FT REPEAT 284 317
 FT REPEAT 318 359
 FT REPEAT 360 391
 FT REPEAT 392 428
 FT REPEAT 429 464
 FT REPEAT 465 488
 FT REPEAT 489 515
 FT REPEAT 516 530
 SQ SEQUENCE 627 AA; 54184 MW; CB9B63779B2C594B CRC64;

Query Match 32.1%; Score 1138.5; DB 1; Length 627;
 Best Local Similarity 47.0%; Pred. No. 2e-44;
 Matches 325; Conservative 36; Mismatches 190; Indels 141; Gaps 36;

23 GGAGAGAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGAAAAAAG---G 78
 DB 2 GGTGPGQ-----GPGGTGP-GGQGPSRPG-----SAAAAAAGAGCGTG 42
 QY 79 AGGAGGGLG-SGGTSGRGLGGGAGAAAAAAGAGAGGAGGGLGSG---GTSGR 133
 DB 43 PGQGPGGYGPQGPGRYGPQGPSPGSAAAAAA--SGGQGPGGYGPQGPGGYGO 101
 QY 134 GGLGGGAGAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGGAGGGLGSGGTSGRGLG----- 187
 DB 102 GQGPSPGSAAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGGAGGGLGSGGTSGRGLG----- 158
 QY 188 -GGAGAGAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGGAGGGLGSGGTSGRGLG----- 242
 DB 159 SGPGSAAAAAAAGAGAGGAGGGLGSGGTSGRGLGGGAGGAGGGLGSGGTSGRGLG----- 217
 QY 243 AGGAGGGLGSGGTSGRGLG---GQ-----GAGAAAAAAGAGAGGAGGGLGSGGTSGRGLG----- 297
 DB 218 QGPBGY-----PGQGPBGYGPQGPGLSGPAGAAAAAAGAGAGGAGGGLGSGGTSGRGLG----- 272


```

01-NOV-1997 (Rel. 35, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE HYPOHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
GN RV2634C OR MYCY441.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeter K., Gas S., Barry C.E., III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajadaram M.A., Rogers J.,
RA Ruter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR HSSP: 280225; CABO2341.1; -.
DR DR
DR HSSP: P04002; IATP.
DR TuberculinList; RV2634C; -.
DR InterPro; IPRO00084; -.
DR InterPro; IPRO01899; -.
DR Pfam; PF00934; Pf; 1.
KW Hypothetical protein.
SQ SEQUENCE 778 AA; 63131 MW; DAB20FE5BE4999E7 CRC64;

Query Match 29.6%; Score 1049.5; DB 1; Length 778;
Best Local Similarity 41.3%; Pred. NO. 1.8e-40;
Matches 291; Conservative 21; Mismatches 294; Indels 99; Gaps 34;

7 GQOMRGSAASRGGLGCGAGCAAAAAAAGACAGCGYGGISG-----GTSGR 57
Db 129 GAPGCGGGLLTGCGNGSGCAPQPGAGAGDGLTGNGTGCGKGDLVSGAAGVGGR 188
Qy 58 GG-TLGCGAGAAAAAAGAGAGCGGYGGLGSOGTSGRGGLGCGAGAAAAAAG 115
Db 189 GGHLLNGGTGAGGAAGATLVG-GTGVYG-CATLLIGSGRG--GAGGAAGAAGVTGG 243
Qy 116 GAGCGGYGGLGSOGTSGRGGLGCGAGAAAAAAGAGAGCGGYGGLGSOGTSGPGGIG 175
Db 244 VGSsgvgvY-FGNGCGFGAGGLGAAGGVCgAAyFTGTGGCGGvGDcAPG--CDGcAG 299
Qy 176 PGGOYTSGRGGLGCGAGAAAAAAGAGAGCGGYGGLGSOGTSGRGGLGCGAGAAAAA 235
Db 300 P--LLTGNNGVGLG-GAGAA-----GNGNGAGAGM-----LLDGCGAGCGGPAVAGY 344
Qy 236 AAAAGAGAGG-----GYGGLGSOGTSGRGGLGCGAGAAAAAAG-----GAGC 283
Db 345 LGGMPPAGAGGGAANMFSGSGAGCGGCTGLAGTNGYNPFSIAMPNTGANCTDSMGNCOT 404
Qy 284 GYGGLGSOGTSGR-GGLGCGAGAAAAAAGAGAGCGGYGGLGSOGTSGPGGYGPQ 342
Db 405 GGNCGGCPAGCVENAGVGGG-----GLGESLDGNDGTGGKGGAG--GRAGRDG-GAG- 455
Qy 343 QTSGRGGLGCGAGAAAAAAGAGAGCGGYGGLGSOGTSGRGGLGCGAGAAAAA 402

```

456 ---GAGCAGGCTGTTDSAGSVATGCEGCGDATTGGVYG--GTGCGAGGKGGGCGHNTGVYDARG 511

QY 403 AAGGAGGGGYGGLGSQGTSGRGGTGGCGACAAAAAAAAGCA--GGGTGGTGLGSQGTSG- 460

Dp 512 GGGGTGGDGCALGAAAG--GNGCGTGGGAGGNGRGKGMILGNGGAGGAGGCTGTGGGAAAG 569

QY 461 RGLGLGGCGACAAAAAAAAGGAAGGAGGCGGGLGSQGTSGPCGYPGCGQTSRGGTGLGCGAG 520

Dp 570 AGGYGAGGEBGLTDGATTAEGGT--GGLGGLG--GYGGTGGMG-----GSGGVGNGNGA 619

QY 521 AAAAAAAAAGGA--GGGYGGLGSQGTSGRGGTGGCG--GACAAAAAAAAGGAAGCGGYG 578

Dp 620 AGSLTGLGGGGGAGCGVGTGGIG--GIGCAGGNGCAGAGACTTTGGGCTTGGCGGTGGVGG 677

QY 579 LGSQGTSGRGGTGGCG--GACAAA-----AAAAAAGCAAGCGTGGTGLGSQGTSGRGGTGGCG 631

Dp 678 AG--GTGGTGGAGGTTGGSSGAGGLIGMAAAGGTGAGGTGGGGGGGLGGG--GNGNGGT 733

QY 632 GAGAAAAAAAAGGAGACGGYG-----LGSQGTSGPGGCTGPGCG 671

Dp 734 GATGGGCGDFALGGNGCAGCAGGSSPAGSSSLGIGMGP-----PGRQ 774

RESULT	7
Y278_MVCTU	
ID	Y278_MVCTU
AC	P56877;
DT	30-MAY-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	HYPOTHEICAL PE-PGRS FAMILY PROTEIN RV0278C. PRECURSOR.
GN	RV0278C OR MTW035.06C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Firmicutes; Actinobacterii; Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCBI_TaxID=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37RV;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA	Hornsbly T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutherford S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA	Taylor K., Whitehead S., Barrell B.G.;
RT	Deciphering the biology of Mycobacterium tuberculosis from the
RT	complete genome sequence.;
RL	Nature 393:537-544(1998).
CC	-1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC	SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: AL021930; CAA17353.1; -
DR	TubercuList; RV0278C; -
DR	InterPro: IPR000084; -
DR	Pfam: PF00934; PE: 1.
KW	Hypothetical protein; Repeat; Signal.
FT	SIGNAL 1 30 POTENTIAL.
FT	CHAIN 31 957 HYPOTHEICAL PE-PGRS FAMILY PROTEIN
FT	RV0278C.
SO	SEQUENCE 957 AA; 81905 MW; 71EBADD417FBA47C CRC64;

Db 578 GCAGGGSTLACGACGAGAGNG-----GTFACGTGGAGSHSTAACTAAGTSCAGCA--G 6253

QY 487 GYGGLGSGQSTGSPGCGTGPBCQQTSGRGGLGCGGAGAAAAAAGAGGCGTGGGLGSGT 5461

Db 626 GDADLLSLGASGAGGSGSGSLTAACTAAGTGGTG----GAGGLLTFGSGGAGSGGFGFNSGN 680

QY 547 SGRGGLGGQ-----GAGAAAAAAGAGGAGGCGGCGGLGSGTSGRGGLGCGGAGAAAA 6010

Db 681 GCAGGAGGCDACLLVSGGAGGAGCAAT--GAATGGDGGAG--GKSGAFGLGGDGGAGGATG 7377

QY 602 AAAAAAGAGGCGGCGGLGSGQSTSGRGGLGCGGAGCA--AAAAAAGAGGAGGCGGCG--LGSQ 6588

Db 738 LSGAFHITGGKGGV--GSAVLIGNGGNGGNGNGKAGKGGAGPFGAGGAGGALLGEN 7959

QY 659 GTSG 662

Db 796 GLNG 799

RESULT	9	ELS_MOUSE	STANDARD:	PRI:	860 AA.
ID	ELS_MOUSE				
AC	P54320;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	ELASTIN PRECURSOR (TROPOELASTIN).				
GN	ELN.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BALB/C; TISSUE=Lung;				
RX	MEDLINE=95130069; PubMed=7829060;				
RA	Wynner K.S., Sechler J.L., Boyd C.D., Passmore H.C.;				
RT	"Use of an Intron Polymorphism to Localize the Tropoelastin Gene to				
RT	mouse chromosome 5 in a region of linkage conservation with human				
RT	chromosome 7.";				
RL	Genomics 23:125-131(1994).				
CC	-1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND				
CC	NOCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.				
CC	-1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER				
CC	INTO AN EXTENSIBLE 3D NETWORK.				
CC	-1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.				
CC	-1- PFM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb.ch).				
CC	or send an email to license@1sb-sib.ch).				
CC	-----				
DR	EMBL; U08210; AAA80155.1; -				
DR	HSSP; P04002; 1WFA.				
DR	MED; MGI:95317; Eln.				
KW	Structural protein; Repeat; Signal; Connective tissue.				
FT	SIGNAL	1	27	POTENTIAL.	
FT	CHAIN	28	860	ELASTIN.	
FT	SEQUENCE	860 AA;	71955 MW;	OC0BE5FAE1EDD7F1 CRC64;	

Query Match	27.5%;	Score 976;	DB 1;	Length 860;
Best Local Similarity	39.38;	Pred. No. 3.1e-37;		
Matches 323;	Conservative 36;	Mismatches 249;	Indels 214;	Gaps 40;

```

0y 13 GSMASGRGLGGG-----AGAAAAAAAAAGAGGGGCG 48
      | : : | | | | | | | | | | | | | | | |
Db 78 GTTACAGPGGTCGAGPCAGLCAFPAGTTPGAGALYPGGAAGAAAYKAAAKACAGLGGCG 137

```

Oy	49	LGSGCTSGRGL--GCGCAGAAAAAAGAGAGGCGYGGJGJSGCTSGRGLGCGAGAAA	107
Dd	138	V--PGVGVGGVPGGVG-----GVPPGCVGVGVPG---GVGGJGGJGGJGVSTGA	184
Oy	108	AAAAAAG-GAG--GGYGGJGJSGCTSGRGLGCGGA-----GAAAAAANA	151
Dd	185	VPPVGGAGTGGAGCRPKGYPGVGLPBGVLPBGTARFPFGVGLPBGVPTGTGKKAKPG	244
Oy	152	AGGAGG--GYGJGJGJSG-----GTSPPGGY-----PQDQTSGRGJGJGQ-----	189
Dd	245	GGGAFSGFPGVPGPGGQPGVPLGPYIPAKRPLPGCYGLPTYNGKLPYGVAGAGKAGPYT	304
Oy	190	--GAGAAAAAAGAGAGGCGY--GLSGCTSGR-----GGJGGGAGAAAAA	237
Dd	305	GTVGSGSQAANAARAKYGAGGAGVLPBGVGGGJPGAGAIPTGJTAGAGTPAAAAAK	364
Oy	238	AAAGGA-----GGGCGYGLSGOSTSGRGLGCGAG	268
Dd	365	AAARAKKGAAGGLVPGGPRVRLPGAGTPTGVGGTIPYVGJLPYGVGPTIGGPTGVGGAV	424
Oy	269	AAAAAAAAAGGA---GGGCGY---GLSGOSTSGRGLGCGAGAAAAAAGAGAGQ	321
Dd	425	SPAAAKAKAAAKYKARGGVGIVTYGVGAGGFPFGYGVGAGJGJGASPAAAAAAKAK	484
Oy	322	GGYGGJGJSGCTSGR--GYPG--QDTSRGGJGCGAGAAAAA-----AG-GAG	370
Dd	485	YGAGGAGJGJLVGVAVPGALPGVAVPVGAGVPGAGTPTAAAAAANAARAKKAGJGPG	544
Oy	371	QGGY--GGJGSGCTSGRGLGJGCGAGAAAAAAGAGAGGCGYGGJGJSGCTSGRGLGJGQ	429
Dd	545	VGGVPGVGVGVIPEGVGVG-----VPGGVPRGGVYTGIG-----ACRPGJGGA	588
Oy	430	GAGAAAAA-----AAAG-GAGGCGYG--GLSGOSTSGRGLGCGAGAAAAA	478
Dd	589	GSPAAKSAAKAAKAAKQYRAAAGJGAGCPGCGAGAGVPGCGAGAGVPGFAGAGVPGFCA	648
Oy	479	AAG-----GAG-----QGGYGGJGJGQ--GTSPPGGYPRGQDQTSRSGJGJGCG--	518
Dd	649	GAGVPGFAGAVPAGSLAARAKYGAAGJGJGCPGJGJGPGJG-----GPGJGJGAGVP	702
Oy	519	---AGAA-AAAAAAGAGAGGCGYGGJGJGJSGCTSGRGLGCGAGAAAAAAGAGAGGCG	574
Dd	703	GRVACAAATPAAAAAARAKAAKAAATQYJGJGAGAGJGAGJGAGJG-----AGJGJGAG	753
Oy	575	GY--GGJGJGCTSGRGLG--GQAGAAAAAAGAGAGGCGYGGJGJGJSGCTSGRGLGJGQ	631
Dd	754	GLGAGJGJGAGJG--GAGJGJGAGAGGVSPAAAAAKAAKYGAAGJGAV--LCARFPGGVAAR	810
Oy	632	GAGAAAAAAGAGAGGCG-----YGG-LGSGCTSGPGYG	667
Dd	811	GGJGSPITPGGGAGJGJGAGVPPRPGAGJGJGJGJGJGJGJGJGJGJGJGJGJGJGJG	852

RESULT	10			
ELS_RAT	ELS_RAT	STANDARD;	PRT;	864 AA.
ID	ELS_RAT			
AC	Q99372;			
DC	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).			
GN	ELN.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91104866; PubMed=1702999;			
RA	Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;			
RT	"Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning."			
RL	Biochemistry 29:9677-9683(1990).			
RN	[2]			


```

Db 297 APCAIPGIGTAGVAGAPDPAAAAAAAAAAKFGAAGLPGVGVPPVGVPGVPGVVP- 355
Oy 310 AAAAAAAAAAGAGGGG--GLGSGGTSGPGGVPGGQOTSGRGGLGQGAGAAAAA- 366
Db 336 -----GAGVPGVGVGVGVPGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV- 407
Oy 367 -GGAGGGGGLGSGGTSGRGGLGSGG-AGAAAAAAGAGAGGGGGLG- 417
Db 408 FGARGAVGIGGIPFGTL-GPGGFGIGDPAAPAAAAAKKAGAGGAGLGVVPGAPG 466
Oy 418 --QGTSRGGLGQGAGAAAAAAG--GAGG-----GCGGLGSGGT- 459
Db 467 AIPGIPGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV- 526
Oy 460 ---GCGGLGSGGAGAAAAA-----GAGG--GGGGLGSGGTSGPG--GYPGQOT 508
Db 527 TGLGCGGVAGVPAAPAAKAAKAPRAAAGLPAVPGGLG--VGAGVPGGLGVAGVPG 585
Oy 509 SGRG-GLGGGGA--GAAMAAAAAAGAGAGGGGGLGSGGTSGRGGLGQGAGAAAAA 565
Db 586 LGVGAAGVPGVAVPGTAAAKAKAFGPGGVGALGGVGDGCGAGITG--GVAAGVPAAPAAA 644
Oy 566 AAGGAGGGGGLGSGGTSGRGGLGCG--GAGAAAAAAGAGAGGGGGLGSGG 621
Db 645 KAAARAAQFGLGSGGGLGGLGAVPGAVGLGVSPAAAKAKAFGAAGLGVLAGDPF 704
Oy 622 TSGRGGLGSGGAGAAAAAAGAGAGGGGGLGSGGTSGPGG 665
Db 705 PLG-GGAGGLG-----VGGKPKPFGALGALGF--PGG 735

```

RESULT 14

```

EBN1_EBV STANDARD: PRT: 641 AA.
AC P03311:
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE EBNA-1 NUCLEAR PROTEIN.
GN BKRF1.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Bignlin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuftnell P.S., Barrett B.G.;
RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RA Nature 310:207-211(1984).
RN [2]
RP SEQUENCE OF 1-26 FROM N.A.
RX MEDLINE=86259739; PubMed=3460083;
RA Sample J., Hummel M., Braun D., Birkenbach M., Kieff E.;
RA "Nucleotide sequences of mRNAs encoding Epstein-Barr virus nuclear
RA proteins: a probable transcriptional initiation site.";
RA Proc. Natl. Acad. Sci. U.S.A. 83:5096-5100(1986).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=90266473; PubMed=2161150;
RA Petit L., Sample C., Kieff E.;
RA "Subnuclear localization and phosphorylation of Epstein-Barr virus
RA latent infection nuclear proteins.";
RA Virology 176:563-574(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 469-607.
RX MEDLINE=96006523; PubMed=7553871;
RA Bocharov A., Barwell J.A., Pfuetzner R.A., Furey W.F. Jr.,
RA Edwards A.M., Strappier L.;
RA "Crystal structure of the DNA-binding domain of the Epstein-Barr
RA

```

```

RT virus origin-binding protein EBNA 1.";
RL Cell 83:38-46(1995).
CC -1- FUNCTION: INVOLVED IN LATENT CYCLE. EBNA-1 FUNCTIONS IN THE
CC MAINTENANCE REPLICATION OF EBV EPISOME. TRANSCRIPTIVATING FACTOR
CC FOR THE ORIGIN AND ENHANCER FUNCTIONS OF ORIP.
CC -1- SUBCELLULAR LOCATION: NUCLEAR, FREE IN THE NUCLEOLUS, SOMEWHAT
CC ASSOCIATED WITH THE CHROMATIN AND HARDLY, IF AT ALL ASSOCIATED
CC WITH THE NUCLEAR MATRIX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; CA24816.1; -.
DR EMBL: M13941; AA45889.1; -.
DR PIR: A03773; Q0BE31.
DR PIR: S33021; S33021.
DR PDB: 1VHI; 23-DEC-96.
DR TRASNSEC: T00211; -.
KW Nuclear protein; DNA-binding; Transcription regulation; Activator;
KW 3D-structure.
FT DOMAIN 87 352 GLY/ALA-RICH.
SQ SEQUENCE 641 AA; 56427 MW; 4D161653E16FC341 CRC64;

```

Query Match 21.5%; Score 763; DB 1; Length 641;
 Best Local Similarity 35.7%; Pred. No. 5.8e-28;
 Matches 232; Conservative 22; Mismatches 209; Indels 186; Gaps 24;

```

Oy 116 GAGGGGGLGSGG--TSGRGGLGAGAGAAAAAAGAGAGGGGGLGSGGTSGPGY 174
Db 5 GPGTGPAGNLCGKCDTSGEGSGSGPQRRGDNHGRGRGRGRGCG--GRGAPGSGSS 62
Oy 175 GPGGQTSR-----GGLGGGAGAAAAAAGAGAGGGGGLGSGGTSGRG 222
Db 63 GPHRGRVRRPQKPSCTICCKTHGGTGAG--AGAGAGAGAGAGAGGAGAG--GAGGAG 119
Oy 223 GLGGGAGAAAAAAGAGAGGGGGLGSGGTSGRGGLGSGGAGAAAAAAGAG 282
Db 120 GAGGAGAG-----GAGAGAG-----GAGAGAGAGAG--GAGAGGGGAGGAG 158
Oy 283 GCGYGLGSGGTSGRGGLGAGAAAAAAGAGAGGGGGLGSGGTSGPGYGPQ 342
Db 159 AGG-----GAGGAGAGAGAG-----GAGGAGAG-----GAGGAGAG-- 187
Oy 343 QTSRGGLGAGAGAAAAAAGAGAGGGGGLGSGGTSGRGGLGSGGAGAAAAA 402
Db 188 --GAGAGAGAGAG-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 235
Oy 403 AAGGAGGGGGLGSGGTSGRGGLGAGAGAAAAAAGAGAGGGGGLGSGGTSGRG 462
Db 236 GAGGAGAG-----AGAGAGAGAGAGAGAGAGAGAGAGAGAGAG-----GAGGAGAG 282
Oy 463 GLGGGAGAAAAAAGAGAGGGGGLGSGGTSGPGYGPQTSRGGLGAGAGAA 522
Db 283 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
Oy 523 AAAAAAAGAGAGGGGGLGSGGTSGRGGLGAGAG--GAAMAAAAAAGAGAGGGGLGS 581
Db 334 -----SGGRGRGGSGGRGSGGRGRGRGRGRGRGRGRGRGRGRGRGRGR 386
Oy 582 QGTS-----GR-----GGLGSGGAGAAAAAAGAGAGGG 613
Db 387 QSSSSGPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPPRRPP 446
Oy 614 YGGLGSGGTSGRGGLG-----GAGAGAAAAAAGAG-----AG 648
Db 447 STGPRGQDGGRRKKGWPKHKGGSNPKENIAGLALLARSHVERTTDEGTWVAG 506

```

```

OY 649 GCGYGLG-----GCGY-----GPGY--GPGOONSGIR 676
DB 507 VFVVGSGTSLYNLRGTAIPQCRTPRLPRLPFGMAPGPGPGLR 555

RESULT 15
ELS_CHICK STANDARD; PRT: 750 AA.
ID ELS_CHICK
AC P07916;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ELASTIN PRECURSOR (TROPOELASTIN) (FRAGMENT).
GN ELN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87242320; PubMed=3593675;
RA Bressan G.M., Argos P., Stanley K.K.;
RT "Repeating structure of chick tropoelastin revealed by complementary
RT DNA cloning."
RL Biochemistry 26:1497-1503(1987).
RN [2]
RP SEQUENCE OF 85-750 FROM N.A.
RX MEDLINE=88309083; PubMed=2841924;
RA Baile V.J., Foster J.A.;
RT "Multiple chick tropoelastin mRNAs."
RL Biochem. Biophys. Res. Commun. 154:1054-1060(1988).
RN [3]
RP SEQUENCE OF 457-750 FROM N.A.
RC TISSUE=Aorta;
RX MEDLINE=87297534; PubMed=3502711;
RA Tokimatsu I., Tajima S., Nishikawa T., Tajima M., Fukasawa T.;
RT "Sequence analysis of elastin cDNA from chick aorta and
RT tissue-specific transcription of the elastin gene in developing chick
RT embryo."
RL Arch. Biochem. Biophys. 256:455-461(1987).
CC -1- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC NCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC -1- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC INTO AN EXTENSIBLE 3D NETWORK.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC -1- ALTERNATIVE PRODUCTS: THERE ARE AT LEAST TWO FORMS OF CHICKEN
CC ELASTIN THAT ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M18633; AAA48761.1; -
DR EMBL: M21880; AAA49082.1; -
DR EMBL: M15889; AAA49108.1; -
DR PIR: A27264; A27264.
DR PIR: A26601; A26601.
DR HSSP: P04002; IMFA.
KW Structural protein; Connective tissue; Repeat; Signal;
KW Alternative splicing.
FT SIGNAL 1
FT CHAIN 25 730 ELASTIN.
FT DOMAIN 83 686 8 X TANDEM REPEATS.
FT REPEAT 83 127 1.
FT REPEAT 219 262 2.
FT REPEAT 263 316 3.

```

```

FT REPEAT 319 393 4.
FT REPEAT 394 482 5.
FT REPEAT 483 554 6.
FT REPEAT 555 619 7.
FT REPEAT 620 686 8.
FT MOD_RES 63 63 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 66 66 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 111 111 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 115 115 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 156 156 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 159 159 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 198 198 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 200 200 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 235 235 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 252 252 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 256 256 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 297 297 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 301 301 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 354 354 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 357 357 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 427 427 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 431 431 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 513 513 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 517 517 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 520 520 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 586 586 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 590 590 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 593 593 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 655 655 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 658 658 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 719 719 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 721 721 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 743 743 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 748 748 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 750 750 OXIDATIVE DEAMINATION (POTENTIAL).
FT MOD_RES 212 212 EMBRYONIC ISOFORM.
FT VARSPLIC 501 501 G -> G (IN REF. 3).
FT CONFLICT 536 536 A -> A (IN REF. 3).
FT CONFLICT 571 571 G -> A (IN REF. 3).
FT CONFLICT 610 610 P -> A (IN REF. 3).
FT CONFLICT 654 654 A -> R (IN REF. 3).
FT CONFLICT 667 667 P -> R (IN REF. 3).
SQ SEQUENCE 750 AA; 63697 MR; E57ECD60C6E556F CMC64.

Query Match 20.0%; Score 709; DB 1; Length 750;
Best Local Similarity 36.5%; Pred. No. 1.5e-25;
Matches 288; Conservative 29; Mismatches 276; Indels 196; Gaps 44;
OY 3 SMTGGQMGKSGM-ASGRGLG-GGAGAGAAAAAGAGGAGGGLGSGTSGRGL 60
DB 30 AIPGGVGGGFFPAGVAGLGGAGLGGKPLKPGVSLGLGLGPGAGVGL 89
OY 61 G-GGGA-----GAAAAA--AAAAGAGGAGGGLGSGTSGRGLG--GGGAGAAA 107
DB 90 GAGLGAFFPGAAFPGAASAAALKAAKAGAGGAGGGLG--GLGGVGGVPGGLVPGV 147
OY 108 AAAAAAG-----GAGGAGGGLGSGT--SGRGLG--GGAAGAAAAAAA 152
DB 148 QPVGAAAKPKPVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 206
OY 153 GGAGGGT-----GGLSGGTSPPGPGQTSRGLGAGGAGAAAAAAG-CA- 205
DB 207 FAGIPGGRPLPFVNL-----GPGIGAGVLAGKAGPYTGAGVGAQAAAAAKKAYAG 260
OY 206 ---GGGAGGGLG---SGTSGRGLGAGGAGAAAAAAGAGGAG--GYGLG-- 253
DB 261 VLPAGAGIPVGGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGVPGV 320
OY 254 -----GTSGRGLGAGG--AGAAAAAAGAGAGGAGG--GLSGGTSGRG 299
DB 321 PGVGVPLVPGVGGIPVAGVGTGAGAAAAAAGAGAGAGAGVPGVPGV 380

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:26 ; Search time 95.41 Seconds
(without alignments)
958.210 Million cell updates/sec

Title: US-09-490-291-6
Predict score: 3543
Sequence: 1 MASTGCGQMGKRGSMASGRG.....TSGIRIRAPSTSFHHHHH 691

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 133305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1998	56.4	617	5	046172 nephila cla
2	1729.5	48.8	544	5	046171 nephila cla
3	1450	40.9	988	5	017434 nephila cla
4	1386	39.1	2639	5	076786 antherea p
5	1260.5	35.6	1489	2	053559 mycobacteri
6	1242	35.1	1079	2	053557 mycobacteri
7	1198.5	33.8	1306	2	053775 mycobacteri
8	1196.5	33.8	1381	2	053552 mycobacteri
9	1180	33.3	1660	2	053215 mycobacteri
10	1167	32.9	1884	5	09nhw2 nephila mad
11	1158.5	32.7	1538	5	053395 mycobacteri
12	1157	32.7	738	5	002402 pinctada fu
13	1153	32.5	2249	5	09nhw4 nephila cla
14	1152.5	32.5	1468	5	09GUB5 gallieria me
15	1152	32.5	1369	2	006810 mycobacteri
16	1151	32.5	853	2	053439 mycobacteri
17	114.5	31.5	749	2	053844 mycobacteri
18	1103	31.1	907	5	044359 nephila cla
19	1084	30.6	871	5	044358 nephila cla

20	1083	30.6	882	2	053845 mycobacteri
21	1038.5	29.3	767	2	053435 mycobacteri
22	1032.5	29.1	837	2	053684 mycobacteri
23	1014	28.6	783	2	053809 mycobacteri
24	1013	28.6	636	5	016987 araneus dia
25	1007	28.4	731	2	050415 mycobacteri
26	992	28.0	714	2	053556 mycobacteri
27	989	27.9	741	2	006808 mycobacteri
28	977.5	27.6	1011	2	006486 mycobacteri
29	975	27.5	810	11	09ES29 mus muscultu
30	957.5	27.0	1729	5	09U617 drosophila
31	937	26.4	923	2	053890 mycobacteri
32	923.5	26.1	667	2	053415 mycobacteri
33	912	25.7	626	5	09nhw1 nephila mad
34	903.5	25.5	694	2	053212 mytilus edu
35	901.5	25.4	904	5	076271 mytilus edu
36	893	25.2	484	2	053394 mycobacteri
37	856.5	24.2	610	5	09V5V8 drosophila
38	848.5	23.9	902	5	016161 mytilus edu
39	848.5	23.9	922	5	044367 mytilus edu
40	833.5	23.5	576	2	P71664 mycobacteri
41	817.5	23.1	584	2	053575 mycobacteri
42	808.5	22.8	618	2	006801 mycobacteri
43	804	22.7	639	2	053952 mycobacteri
44	801	22.6	606	2	053884 mycobacteri
45	800	22.6	591	2	007224 mycobacteri

ALIGNMENTS

```
RESULT 1
ID 046172 PRELIMINARY; PRT; 617 AA.
AC 046172:
AD 01-JUN-1998 (TREMUREL.06, Last sequence update)
DT 01-JUN-1998 (TREMUREL.06, Last sequence update)
DT 01-MAR-2001 (TREMUREL.16, Last annotation update)
DE DRAGLINE SILK PROTEIN SPIDROIN 1 (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165058; PubMed=8120021;
RA Beckwith R., Arcidiacono S.;
RT "sequence conservation in the C-terminal region of spider silk
RT proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus
RT bicentenarius (Araneidae).";
RL J. Biol. Chem. 269:6661-6663(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Beckwith R., Arcidiacono S., Stote R.;
RL Insect Biochem. Mol. Biol. 0:0-0(1998).
DR EMBL; U37520; AAC04504.1; -.
FT NON_TER 1 1
FT NON_TER 617 617
SQ SEQUENCE 617 AA; 49410 MW; 1EFOCE92269A832E2 CRC64;
```

Query Match 56.4% Score 1998; DB 5; Length 617;
Best local similarity 66.6%; Pred. No. 5, se-123;
Matches 462; Conservative 18; Mismatches 58; Indels 156; Gaps 30;

```
QY 2 ASMTGCGQMGKRGSMASGRGIGCGGAGAAAAAGAGAGCGGCGTSGTSGRGIG 61
DB 3 AAAGAGAGCGGIGGIGCGAGCGGAGAGAGAGAGAGCGGCGGAGAGAGAGAG 57
QY 62 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
DB 58 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 100
```

Qy	122	YGGLSGSGTSGRGLGSGGAGAAAAAAGGAGCGGCGYGGGLSGSGTSGPGGCGPQDQTS	181
Db	101	YGGGLSGG-AGRGSGSGGQAG-AAAAGAGAGGGGCGYGGGLSGG-----A	142
Qy	182	GRGGLGCGAGAGAAAAAAGAG-----GAGCGGCGGLSGSGTSGRGLGCGGAGA	231
Db	143	GRGLGGGCGAGAAAAAAGCAGCGCGYGGGLGCGAGCGGCGYGGGLSGG-AGRGGLGSGGAG-	200
Qy	232	AAAAAAGAGCGGCGYGGGLSGSGTSGRGLGCGGAGAAAAAAGAGAGCGGCGGLG	291
Db	201	-----AAAAAAGAGAG-----GGLGCGGAGCGAGAAAAAAGCAGCGGCGGLG	244
Qy	292	CGTSGRGLGCGGAGAAAAAAGCAGCGCGYGGGLSGSGTSGPGCGYGP-GQGTSGRGL	350
Db	245	CG-----AGRGCGAGAAAAAAGV-----GAGCGGCGGCGG-AGCGGCGGLSGCGAGRGL	292
Qy	351	GGCGAGAAAAAAGAGAGCGGCGYGGGLSGSGTSGRGLGCGGAGAAAAAAGAGAGCGG	410
Db	293	GGCGAG-----AAAAAAGAGAG-----GGLGCGGAGCGAGAAAAAAGAGAGCGG	335
Qy	411	GYGGLSGSGTSGRGLGCGGAGAAAAAAGAGAGCGGCGYGGGLSGSGTSGRGLGCGGAG	470
Db	336	GYGGLGNGG-AGRGCGG-----AAAAAAGAGCGGCGYGGGLSGG-AGRGGLGCGGAG	384
Qy	471	AAAAAAGAGAGCGCGCGYGGGLSGSGTSGPGYGP-GQGTSGRGLGCGGAGAAAAAAG	529
Db	385	-----AAAAAAGAGCGCGGCGYGGGLSGG-AGCGGCGGLGSGSGRGLGCGGAG-AAAA	435
Qy	530	AGGAGCGGCGYGGGLSGSGTSGRGLGCGGAGAAAAAAGAGAGCGGCGYGGGLSGSGCG	588
Db	436	AGGAGG-----GGLGCGGAGCGAGAAAAAAGCGYRGGCGYGGGLSGG-----AG	477
Qy	590	LGCGGAGAAAAAAGAGAGCGGCGYGGGLSGSGTSGRGLGCGGAGAAAAAAGAGAGCG	648
Db	478	RGGCGAG-----AAAAAAGAGAGCGGCGYGGGLGCGG-GRGGLGCGGAG-----AAAA	526
Qy	650	GGYGGGLSGSGTSGPGCGYGPQGTSGIRARASTS	683
Db	527	GGYGGGSGAGAA-----SMAASRLSPDAS	552
RESULT	2		
ID	046171	PRELIMINARY;	PRT; 544 AA.
AC	046171;		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	SPIDROIN 1 (FRAGMENT).		
OS	Nephila clavipes (Orb. spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Eutelegynae; Araneoldea; Tetragnathidae; Nephila.		
OC	NCBI_TaxId=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90384959; PubMed=2402194;		
RA	Xu M., Lewis R.V.;		
RT	"Structure of a protein superfiber: spider dragline silk."		
RL	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).		
RN	[2]		
RP	SEQUENCE OF 449-544 FROM N.A.		
RX	MEDLINE=98148687; PubMed=9487707;		
RA	Arcidiano S., Mello C., Kaplan D., Cheley S., Bayley H.;		
RT	"Purification and characterization of recombinant spider silk		
RT	expressed in Escherichia coli."		
RL	Appl. Microbiol. Biotechnol. 49:31-38(1998).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Beckwith R., Arcidiano S., Stote R.;		
RL	Insect Biochem. Mol. Biol. 0:0-0(1998).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RP	Beckwith R.;		

RL	Submitted (JAN-1995) to the EMBL/GenBank/DDBJ databases.
FR	EMLBL; U20329; AAC38957.1; -.
FT	NON_TER 1 1
SQ	SEQUENCE 544 AA; 44107 MW; CC611B351945615 CRC64;

Query Match 48.8% ; Score 1729.5; DB 5; Length 544;
Best Local Similarity 61.3%; Pred. No.1,4e-105;
Matches 409; Conservative 12; Mismatches 33; Indels 213; Gaps 28;

```

OY      18 GRGGLGGCGAG-AAAAAAAAAAAGAAGCAGGCATGCTGTSGRKGTLGGCAAAAAAAA 76
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       1 GYGGLGGCGAGAGAAAAAAAGGAAGCAGCATGCTGTSG---AGRRCGAC-----AAAAA 52

OY      77 GGAGCGTGTGCLGSOGTSRGKLTGGCAGAAAAAAAAGAACAGCGTGCLGSOGTSGRGL 136
           ||||| | : | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       53 GGAGCGTGTGCLGSOG--AGRGLTGGCAG----AAAAAGGVQG-----GGI 92

OY      137 GGCGAGAAAAAAGGAAGCAGCGTGTGCLGSOGTSPGCVGPQQOTSGRKLGGCGAGAAAA 196
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       93 GGCGAGCGAGAAAAAAGGAAGCAGCGTGTGCLGSOG-----AGRGSSGGCAG---- 135

OY      197 AAAAAAGCAGCGTGTGCLGSOGTSGRKLTGGCAGAAAAAAAAGAACAGCGTGTGCLGSOGT 256
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       136 AAAAAAGCAGCGTGTGCLGSOG--AGRGLTGGCAG---AAAAAAGCACGCGTGTGCLGQ-- 189

OY      257 SGRGLGGCGAGAAAAAAGCAAAGCAGCGTGCLGSOGTSGRKLTGGCAGAAAAAAA 316
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       190 -----GAGCGTGTGCLGSOG--AGRGLTGGCAG-----AAAAA 220

OY      317 GGAGCGTGTGCLGSOGTSRPVGYVPQQOTSRRKLTGGCGAGAAAAAAAAGAACAGCGTYG 376
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       221 GGAGQ-----GILGGCAG-----AAAAAAGCACGCGTYG 250

OY      377 LGSOGTSGRKLTGGCGAGAAAAAAGAACAGCGTGCLGSOGTSGRGLGGCGAGAAAA 436
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       251 LGSQ-----AGNAGCGAG-----AAAAAGCAGCGTYG----- 280

OY      437 AAAAAAGCAGCGTGTGCLGSOGTSGRKLTGGCAGAAAAAAAAGAACAGCGTGTGCLSOQT 496
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       281 -----QCAGCGTGTGCLGSOG--AGRGLTGGCAG---AAAAAAGCACAGQ----- 320

OY      497 SGPGVTGPQQOTSGRKLTGGCGAGAAAAAAGAACAGCGTGCLGSOGTSGRKLTGGCG 556
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       321 -----GELGGCAG-----AAAAAGCAGCGTGTGCGCA-----GGG 351

OY      557 AGAAAAAAGAACAGCAGCGTGCLGSOGTSGRKLTGGCGAGAAAAAAAAGAACAGCGCYG 616
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       352 ACAAAAAAGAACAGTCAGCGTGCLGSOG-----AGRGGCAG-----AAAAAAGCACGCGCYG 403

OY      617 TCSOGTSGRGLTGGCAGAAAAAAGAACAGCAGCGTGCLGSOGTSPGTVGPQQOTSGIR 676
           ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       404 TGCGGT-GRGLTGGCGAGAAAAAV-----GACGCGTGTGVSASNA-----SMAAR 448

OY      677 TRAPSTS 683
           . . | |
Db       449 LSPPAS 455

```

RESULT 3
ID 017434 PRELIMINARY: PRT: 988 AA.
AC 017434:
DT 01-JAN-1998 (TREMBLrel., 05, Created)
DR 01-JAN-1998 (TREMBLrel., 05, Last sequence update)
DI 01-OCT-2000 (TREMBLrel., 15, last annotation ampule)
DE MINOR AMPLULATE SLIK PROTEIN MSPI1 (FRAGMENT).
OS Nephila clavipes (Ord spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
XC Araneoidea; Entelegynae; Aranidae; Tetragrathidae; Nephila.
OX NCBI_TaxId=6915;
RN [1]
RP SEQUENCE FROM N.A.


```
Db 1348 SAAAAAAAAAASGAGSGGCGYGCYSDSAAAAAAAAASGAGSGCYGC 1407
QY 579 LGSQCT-----SGRGGLG-----OGAGAAAAAAAAAGAGAG 612
Db 1408 YGSDSAAAAAAAAASGAGYGVGMODGYGSDSAAAAAAAAASGAGSG 1467
QY 613 GYCGIGS-----OCTSGRGLGCGAGAAAAAAAA-----AGAG 648
Db 1468 GYCGYSDSAAAAAAAAASGAGAGGYGSDSAAAAAAAAASGAGAG 1527
QY 649 GCGY-----GGLG-----OCTSGPGGPGPGQTSGRIRAPSTFEHH 687
Db 1528 -GGYWGCGGCGYSAAAAAAAAAASGAGRGDGGYGSSSAAAAAAAAARRAG 1586
QY 688 H 688
Db 1587 H 1587

RESULT 5
053559 PRELIMINARY; PRT; 1489 AA.
AC 053559:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PGSR-FAMILY PROTEIN.
GN RV3514 OR MTW023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kiroh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL022022; CAAL7751.1; -.
DR TubercuList; RV3514; -.
DR InterPro; IPR000084; -.
DR InterPro; IPR002173; -.
DR Pfam; PF00934; PE; 1.
DR ProDom; PD001223; -; 1.
DR PROSITE; PS00583; PKR_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;

Query Match 35.6%; Score 1260.5; DB 2; Length 1489;
Best Local Similarity 38.5%; Pred. No. 9.2e-75;
Matches 344; Conservative 35; Mismatches 274; Indels 241; Gaps 37;
```

```
QY 132 GRGIGGCGA-----GAAAAAAAAAGGA-----GQGYGGLGSGQTSRPGY 175
Db 392 GDGGAGGAGADADPGATGCTGTFAGAGAGAGSSGAGGTNGSGGCGAGAGGA 451
QY 176 PGQOTSGRGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGQTSRGRGLGQ----- 227
Db 432 GADNFTGIGTGDBGCGTGGAAGAGAGGAAGTGTGCMIGTTGNMAGYVAGAGGQGGDGGAG 511
QY 228 GAGAAA-----AAAAAAGGA-----GQGYGGLGSGQ----- 255
Db 512 GAGADADPGATGCTGTFAGAGAGAGAGSSGAGGTNGSGGAGCTGGCGAGAGAGAD 571
QY 256 -TSGRGGLGCGAGAAAAAAAAAAGAGAGCGYGC-LGSGQTSRGRGLGQ-----GAGAAA 309
Db 572 NPTGIGTGCGDGTGTGAAGAGAGAGAACTGTGTGMICTTNMAGYVAGAGGQGGDGGAGG 631
QY 310 AAA-----AAAGAGAGQ-----GGYGLGSGQ-GTSGPGYPGQOT 344
Db 632 ADADPGATGCTGTFAGAGAGAGAGSSGAGGTNGSGAGCTGRSGAGAGADNPT 691
QY 345 SGRGGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGQTSRGRGLGCGAGAAAAAAAA 403
Db 692 -GIGGTGDDGTGGAAGAGAGAGAGTGTGCMIGTTGNMAGYVAGAGAGAGAGAGAG 750
QY 404 AGAGAGCGYGGLGSGQCT---SGRGLGCGAGAAAAAAAAAAGAGGCGYGC-LGSGQTS 459
Db 751 GGTDDGGAGAGAGAGADNPTGTGTGTGDDGTGTGAAGAGAGAGAGAGTGTGCMIGTTGNA 810
QY 460 GRGGLGCGQ-----GAGAAAAA-----AAAAAGGA----- 483
Db 811 GYGGAGCGCGGCGAGAGAGADADPGATGCTGTFAGAGAGAGAGSSGAGCTNGSGAGGT 870
QY 484 -----GQGYGGLGSGQTSRPGY-YRPGQOTSGRGLGCGAGAAA 523
Db 871 CGQVYVAGAGAGISFNSNGSGCTGTGTGCGY--GTGDDGAGNAGTACDPKXGTGTGTG 927
QY 524 AAAAAAAAAAGAG-OGCYGGLGSGTSGRGLGCGAGAAAAAAAAAAGAGGCGYGGLGSG 582
Db 928 SCGAGGCGANFNNGTGTGTG--GTGKGGLMTDGLSSATISGTGTGTGCGKGTGAGCD 985
QY 583 GTSRGLGCGAGAAAAAAAAAAGGA-----GQGYGGLGSGQTSRGRGLGCGQ----- 631
Db 986 SAGTGTGTGTGAGAGNAGAGLANTGTAGNAGTGTGCGGCGAGGCGGDSG-SGLGCGPFRAG 1044
QY 632 -----GAGAAAAA-----AAAAAGAGAGCGYGLG-----SGQTSRPGY 667
Db 1045 GAGGKAGAGSSGAGGTNGSGAGAGAGGCGAGAGAGISFNSNGCTGTGTGCGV 1098

RESULT 6
053557 PRELIMINARY; PRT; 1079 AA.
AC 053557:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PGSR-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTW023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Kiroh A., Mclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
```

RA Taylor K., Whitehead S., Barrell B.G.:
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 Nature 393:537-544(1998).
 RL Nature 393:537-544(1998).
 DR EMBL: AL022022; CAA17749.1; --
 DR TubercuList; Rv3512; --
 DR InterPro: IPR002202; --
 DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
 FT NON_TER 1
 SQ SEQUENCE 1079 AA; 81163 MW; A79718CDB74B97D CRC64;

Query Match 35.1%; Score 1242; DB 2; Length 1079;
 Best Local Similarity 40.7%; Pred. No. 1,1e-73;
 Matches 316; Conservative 34; Mismatches 306; Indels 120; Gaps 29;

OY 5 TGGGOMRGSMASGRGLGCGAGAAAAAAGA-GGGGCGG-----LGSQG 53
 DB 58 TGGGCGNGCN--GGNGGTGKGTGCGDCLAGSSGAGKCGKCGDAGKAGTGSAPGTAG 115 -
 OY 54 TSGRGLGCGG-----AGAAAAAAGA-GGGGTGCGTSGRGLGCG- 101
 DB 116 TGGDGGKGGMGCGTGAAGTTGPGVGTGASGCTGCGSGAGGTGGDGAANGCTAGAGAGAGNG 175
 OY 102 -----GAGAAAAAAGAAGAGGCGGCG-----LGSQTSRGLGCGGAGAAA 145
 DB 176 GKGGDGAGACTSTTAGNSGAGSGGKGGDAGAGAGACTPGANGTAGNGDGD--GAG 233
 OY 146 AAAAAAGAGGCGGCGGCGG-----GTSCPGCGPGQOTSGRGLGCGAGAAAAA 201
 DB 234 AVGISGATGAGDGGHGTGAAGNGGTGAGGSGSIDVGCGGTGCGGNGGALGAGAGD 293
 OY 202 AGCAG--GGGTGCGG-----GTSGRGLGCGG--AGAAAAAAGAAGAG 244
 DB 294 AGCGSGNGGNGTGGKGGNAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 353
 OY 245 GGGGTGCGTSGRGLGCGG-----AGAAAAAAGAAGAGAG-----GGG 286
 DB 354 TGSAGGNG--GTGGRGSGGAGGCGTGGGCGKGGNAGDGVGGAGGSGPPTSPEGN 411
 OY 287 GGLSGQTSRGRGLGCGGAGAAAAAAGAAGAGCGGCGGCGGCGG-----GGY 343
 DB 412 GCGGCGGSGGAG--GAGAGAGAGGAGNGTNGNGCGGCGTGGAGAAASATNGSGGAGG 470
 OY 344 TSGRGLGCGGAGAAAAAAGAAGAGCGGCGGCGGCGGCGG-----A 393
 DB 471 TGGCGSGGAGGTGCGAGTGGAGAGDGGCGCGGAGG--GAGCGGAGGAGTGGNGNITG 529
 OY 394 GAAAAAAGAAG--AGGGGCGGCGG-----GTSGRGLGCGGAGAAAAAAGAAGAGAG 450
 DB 530 GTACTAGAGAGAGAGAGAGGCGGCGGCGGCGGAGGAGGAGGCGDRTVGGGTVPAGS 589
 OY 451 GGLSGQTSRGRGLGCGGAGAAAAAAGAAGAGCGGCGGCGG-----SQGTSGPGY 505
 DB 590 GCGG--GNAGGAGAGCG--GGADGSGGCGDAGTGGGNGGNGNRSNGTGGAGAGNGG 645
 OY 506 GQTSRGRGLGCGGAGAAA--AAAAAAGAGCGGCGG--GCGGCGTSGRGLGCGGAGAAAA 563
 DB 646 GAGCAGAGAGAGCGGCGTGGCGAGCGDAGCGNGGNGGNGGNGGNGGNGGNGGNGGNGG 705
 OY 564 AAAAAAGAGCGGCGG-----GSGGT--SGRGLGCGGAGAAAAAAGAAGAGAGCGG 618
 DB 706 GSGNGGNGSGGNGAGNGGNGSGTGGSDGGAGAGAGAAAGTGGTGGDGLTGTGGTGGSG 765
 OY 619 SQGTSGRGLGCGGAGAAAAAAGAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 674
 DB 766 --GTGGCGGNGAGADYANMTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 819

DT 01-JUN-1998 (TREMREL. 06, Created)
 DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
 DT 01-OCT-2000 (TREMREL. 15, Last annotation update)
 DE PGSS-FAMILY PROTEIN.
 GN Rv0578C OR MT039.16C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID:1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:H37RV;
 RX MEDLINE:98295987; PubMed:9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Kiroh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajadaram M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.:
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021942; CAA17449.1; --
 DR TubercuList; Rv0578c; --
 DR InterPro: IPR000084; --
 DR InterPro: IPR000209; --
 DR InterPro: IPR002173; --
 DR Pfam: PF00934; PE; 1.
 DR ProDom: PD001223; --; 1.
 DR PROSITE: PS00583; PKK_KINASES_1; UNKNOWN_2.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
 SQ SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match 33.8%; Score 1198.5; DB 2; Length 1306;
 Best Local Similarity 38.4%; Pred. No. 8.8e-71;
 Matches 323; Conservative 37; Mismatches 293; Indels 189; Gaps 34;

OY 11 GRGSMASGRGLGCGGAGAAAAAAGAAG-----AGGGGCGGCGGCGGCGGCGG 66
 DB 399 GTTSTSGNGDGGGKAGADAISSGOTGANGRGCGDGGGCGGCGGAG--GAGGCGAGAGLFG 456
 OY 67 AAAAAAAGAAG--AGGGGCGG--LGSQTSRGRGLGCGG-----GAAAAA 114
 DB 457 SEAPRPGAGGTGAGAGNGGTQAGDGGTGGAGAGAGGCGGAGAGAGAGAGAGAGAGAG 516
 OY 115 GG-----AGGGGCGG-----GSGGTSGRGLGCGGAG-----GAAA 146
 DB 517 GNGGNGGPGGAGGAGGAGGAGLALAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 576
 OY 147 --AAAAAAGAGGCGGCGGCGGCGG--PGGCG--PGQ--QTSRGLGCGGAGAAAAA 199
 DB 577 NGVYAGAGAGHGDPEVVGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 636
 OY 200 AAAG-----AGGGGCGGCGG-----GTSGRGLGCGGAGAAAA 234
 DB 637 GASGGRGCGDGLVNGGAG 696
 OY 235 AAAAAAG--AGGGGCGGCGG-----GTSGRGLGCGGAG-----GAAAAA 277
 DB 697 GGTGAGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 756
 OY 278 AGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 335
 DB 757 AGSGGPGAGGCG--GTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 814
 OY 336 GCGP-----GQTSRGRGLGCGG-----AGAAAAA 364
 DB 815 GGAGPAGTTGCGAG 874
 OY 365 AAGG-----GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407

```
Db 875 GGGAGSTGATGSSSSIHVNGNGNGNGDHALLSNGAAGNGNGSLRGSGAG 934
Oy 408 GGGGCGTGGTSGRGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 460
Db 935 GGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 994
Oy 461 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 509
Db 995 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1052
Oy 510 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 558
Db 1053 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1110
Oy 559 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 618
Db 1111 TTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1168
Oy 619 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 672
Db 1169 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1227
Oy 673 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
Db 1228 AG 1229

RESULT 8
ID 053552 PRELIMINARY; PRT: 1381 AA.
AC 053552;
DT 01-JUN-1998 (TREMBLREL. 06, Created)
DT 01-JUN-1998 (TREMBLREL. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
DE GRS-FAMILY PROTEIN.
GN RV3507 OR MTW023.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekela F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagers K., Kiroh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellern S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAAL1744.1; -.
DR HSSP: P00778; ZULL.
DR Tuberculist: RV3507; -.
DR InterPro: IPR000084; -.
DR InterPro: IPR002173; -.
DR Pfam: PF00934; PE: 1.
DR ProDom: PD001223; ?; 1.
DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1381 AA; 110624 MW; CA09676BD07F6482 CRC64;
```

```
Db 150 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 209
Oy 56 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 105
Db 210 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 269
Oy 106 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 161
Db 270 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 325
Oy 162 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 218
Db 326 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366
Oy 219 -SGRGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 276
Db 367 AAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
Oy 277 AAG---GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 303
Db 427 AAGDPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 486
Oy 304 ---GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 339
Db 487 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 546
Oy 340 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
Db 547 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 598
Oy 400 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 449
Db 599 TAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 648
Oy 450 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
Db 649 KAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708
Oy 499 -PG-GPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 544
Db 709 TTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 768
Oy 545 ---GTSRGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584
Db 769 GAGPHADAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 828
Oy 585 S---GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 640
Db 829 TTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 884
Oy 641 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683
Db 885 AAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 927

RESULT 9
ID 053215 PRELIMINARY; PRT: 1660 AA.
AC 053215;
DT 01-JUN-1998 (TREMBLREL. 06, Created)
DT 01-JUN-1998 (TREMBLREL. 06, Last sequence update)
DT 01-OCT-2000 (TREMBLREL. 15, Last annotation update)
DE GRS-FAMILY.
GN RV2490C OR MTW008.46C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
```

RA Gordon S.V., Eiglmeler K., Gas S., Barry C.E. III, Tekela F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osbourne J., Quail M.A., Rajadream M.A., Rogers J.,
 RA Rulter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL021246; CAAL6067.1; -.
 DR TubercuList: RV2490c; -.
 DR InterPro: IPR000084; -.
 DR InterPro: IPR000228; -.
 DR InterPro: IPR002173; -.
 DR Pfam: PF00934; PE: 1.
 DR PRODOM: PD001223; -; 1.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
 DR PROSITE: PS01287; RTC; UNKNOWN_1.
 SO SEQUENCE 1660 AA; 133124 MW; 3A89CE12C0FA945 CRC64;

Query Match 33.3%; Score 1180; DB 2; Length 1660;
 Best Local Similarity 37.6%; Pred. No. 1.7e-69;
 Matches 326; Conservative 38; Mismatches 292; Indels 212; Gaps 35;

QY 2 ASWTGCGQMGKSMASRGGLGCGAGAAAAAAGAG-----GGG 45
 DB 514 SSGTPECDGNGC-ACGACAGCAHAGDCAGAGGNGAGAGCAHFNALVSDG 572
 QY 46 YGGLSGQTSRGGLGCGAGAAAAAAGAGGCGYGGCL-GSGQTSRGGLGCGAG 104
 DB 573 NGDGG--GAGGKGGDGGAG-GAGGDPAPRAGSGVGGDGGAGAGAPNGSGGGGDM 629
 QY 105 A---AAAAAAGAGAGCGCGYGGGLGSGQ-----TSGRGGLGCGAGAAAAA 147
 DB 630 AFKPDGCGAGGCGDGPAGGKGGAGAGATGEGTGAATVHSGNGKGNADATVAG 689
 QY 148 AAAAAGCA-----GCGYGGGLGSGQTSGPGGYGPGQ-----TSGRG 184
 DB 690 ANCGKAGACNGCLVGDGAGGDCGSGAAGANGANYCEDADTLGSGPGESEANGCG 749
 QY 185 GCGGAGAGAAAAAAGAGAGCGYGGGLGSGQTSRGGLGCGAGAAAAAAGAGAG 244
 DB 750 GYGGGAGAGAGCGDGGAGSSALGSGNGGRDAGAG---GAGCAGAGAGAGSVSGDG 805
 QY 245 QCGYGGGLGSGQTSRGGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 303
 DB 806 PGKGGAGAGAGAGAGAGAG--GGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 862
 QY 304 G--AGAAAAAAGAGAGAGCGYGGGL-GSQGTSGPGGYG-----PGQTSGRG 349
 DB 863 GAG 920
 QY 350 LGG-AGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 387
 DB 921 LGGDSGNGTAAASGVDAASHGPGSGGNGGNGAGAAVAGAGAGAGAGAGAGAGAG 980
 QY 388 LCGGCGAGAAAAAAG 437
 DB 981 AGGNGAG 1040
 QY 438 AAAAAG 488
 DB 1041 DGGAG 1100
 QY 489 GG-----LGSQGTSGPGGYG-----PGQTS 509
 DB 1101 GGAAPSGTGVSHGTGCGVGGGLGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
 QY 510 GRCGLGAG-----GAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 545
 DB 1159 GRCGLGAG 1216

QY 546 TSGRGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 595
 DB 1217 LTFDDGAGGAGAGTAEEALGSDGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1276
 QY 596 -----GAAAAAAG 646
 DB 1277 VAGGNGAG 1336
 QY 647 AGGCGYGGGLGSGQTSGPGGYGPGQTS 674
 DB 1337 NGGCGTGGCGGAG 1364

RESULT 10
 Q9NHW2
 ID Q9NHW2 PRELIMINARY; PRT; 1884 AA.
 AC Q9NHW2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
 GN FLAG.
 OS Nephila madagascariensis.
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
 OX NCBI_TaxId=115969;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20156766; PubMed=10688794;
 RA Hayashi C.Y., Lewis R.V.;
 RT "Molecular architecture and evolution of a modular spider silk protein
 gene."
 RL Science 287:1477-1479(2000).
 DR EMBL: AF218623; AAF36091.1; -.
 DR InterPro: IPR00087; -.
 FT NON_TER 1884 1884
 SO SEQUENCE 1884 AA; 148738 MW; 27B6F45339PD20A5 CRC64;

Query Match 32.9%; Score 1167; DB 5; Length 1884;
 Best Local Similarity 40.0%; Pred. No. 1.3e-68;
 Matches 318; Conservative 36; Mismatches 295; Indels 146; Gaps 36;

QY 6 GCGQMGKSMASRGGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 64
 DB 559 GGS--GCGGGLGIRSGCGGAGVPG--GSGPGSICPGSSGCGGLGPGSGGGLGPGSGSPG 615
 QY 65 AGAAAAAAG 115
 DB 616 VSGGCGVGPYGPAGSGGPGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 675
 QY 116 GAGGCGYGGGLGSGQTSRGGLGCGAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAG 168
 DB 676 GAG-GPYGPAGPGYGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 734
 QY 169 SGPYGPAGQTSRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 225
 DB 735 AGPGYGPAG--GAGPGYGPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 791
 QY 226 GCGAG-AAAAAAG 282
 DB 792 PGAG 851
 QY 283 QGGYGGGLGSGQ-----GTSGRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 316
 DB 852 SGGAGSGGTTTIEDDITVDANGPPTTISEELTIGAGAGAGAGAGAGAGAGAGAGAG 911
 QY 317 GAG 373
 DB 912 GGVGPAGSGPGAG 968
 QY 374 YGGLGSGQTSRGGLGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:58:11 ; Search time 86.32 Seconds
(without alignments)
370.823 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700
Sequence: 1 MASMTGGCGMGRIRIRGYCG.....GLSGCDVLQALLGHHHHH 528

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/AA1980.DAT:*
2: /SID58/gcgdata/geneseq/AA1981.DAT:*
3: /SID58/gcgdata/geneseq/AA1982.DAT:*
4: /SID58/gcgdata/geneseq/AA1983.DAT:*
5: /SID58/gcgdata/geneseq/AA1984.DAT:*
6: /SID58/gcgdata/geneseq/AA1985.DAT:*
7: /SID58/gcgdata/geneseq/AA1986.DAT:*
8: /SID58/gcgdata/geneseq/AA1987.DAT:*
9: /SID58/gcgdata/geneseq/AA1988.DAT:*
10: /SID58/gcgdata/geneseq/AA1989.DAT:*
11: /SID58/gcgdata/geneseq/AA1990.DAT:*
12: /SID58/gcgdata/geneseq/AA1991.DAT:*
13: /SID58/gcgdata/geneseq/AA1992.DAT:*
14: /SID58/gcgdata/geneseq/AA1993.DAT:*
15: /SID58/gcgdata/geneseq/AA1994.DAT:*
16: /SID58/gcgdata/geneseq/AA1995.DAT:*
17: /SID58/gcgdata/geneseq/AA1996.DAT:*
18: /SID58/gcgdata/geneseq/AA1997.DAT:*
19: /SID58/gcgdata/geneseq/AA1998.DAT:*
20: /SID58/gcgdata/geneseq/AA1999.DAT:*
21: /SID58/gcgdata/geneseq/AA2000.DAT:*
22: /SID58/gcgdata/geneseq/AA2001.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	2319	85.9	646	18	AAW27178
2	2196	81.3	718	19	AAW53346
3	2196	81.3	718	21	AAV59070
4	2182	80.8	718	12	AAW4308
5	1999	74.0	651	20	AAV40097
6	1883.5	69.8	606	16	AAW9035
7	1883.5	69.8	606	20	AAV40102
8	1879.5	69.6	604	16	AAW9057
9	1871.5	69.3	604	16	AAW9053
10	1858.5	68.8	606	20	AAV40100
11	1858.5	68.8	606	20	AAV40100

12	1344	49.8	831	16	AAW80168
13	1234	45.7	615	20	AAV40099
14	1144	42.4	595	12	AAW4309
15	1144	42.4	595	19	AAW53347
16	1144	42.4	595	21	AAV59071
17	973.5	36.1	465	22	AAW70188
18	972	36.0	1177	9	AAW80940
19	972	36.0	1177	11	AAW80307
20	972	36.0	1177	17	AAW85105
21	972	36.0	1177	18	AAW26342
22	972	36.0	1177	21	AAW78277
23	972	36.0	1177	22	AAW2725
24	972	36.0	1177	22	AAW63995
25	972	36.0	1178	14	AAW41007
26	972	36.0	1178	19	AAW53518
27	970	35.9	714	16	AAW9059
28	970	35.9	714	20	AAV40103
29	962.5	35.6	1023	19	AAW53524
30	962.5	35.6	1059	14	AAW41013
31	962.5	35.6	1059	18	AAW26348
32	962.5	35.6	1101	21	AAW78283
33	955.5	35.4	1059	9	AAW82962
34	946.5	35.1	980	16	AAW81318
35	929	34.4	1332	17	AAW95109
36	929	34.4	1332	22	AAW2729
37	929	34.4	1332	22	AAW63999
38	922.5	34.2	1038	17	AAW95107
39	922.5	34.2	1038	22	AAW2727
40	922.5	34.2	1038	22	AAW63997
41	921.5	34.1	695	19	AAW53526
42	921.5	34.1	766	18	AAW26349
43	921.5	34.1	766	21	AAW78285
44	921.5	34.1	979	21	AAW78286
45	921.5	34.1	1050	18	AAW26350

ALIGNMENTS

RESULT 1	AAW27178	standard; Protein: 646 AA.
AC	AAW27178;	
DT	09-DEC-1997	(first entry)
DE	Nephila clavipes spider silk protein.	
KW	High strength film; fibre; woven article; parachutes; sails; absorber; body armour; heavy metal; biological weapon; chemical; flavour; fragrance; Nephila clavipes.	
OS	Nephila clavipes.	
PN	W09708315-A1.	
PD	06-MAR-1997.	
PE	22-AUG-1996;	96WO-US13767.
PR	22-AUG-1995;	95US-0517694.
PA	(BASE/) BASEL R M.	
PA	(ELIO/) ELION G R.	
PI	Basel RM, Elion GR;	
XX	WT; 1997-179272/16.	
DR	N-PSDB; AAT85356.	
XX	New opt. multimerised DNA sequences encoding spider silk protein - contg. both repetitive and non-repetitive sequences, useful for	

PT making high strength films, fibres, woven articles etc.

PS Example 2; Flg 1; 57pp; English.

A process has been developed for the production of a DNA fragment encoding silk protein. The process involves: (a) selecting target DNA, from a silk-producing spider, that contains many repetitive and non-repetitive regions; (b) selecting a single-stranded DNA primer of at least 10 nucleotides with a sequence that is complementary to a region of the target; (c) repetitively combining the primer with melted target DNA, incubating the mixture with nucleotides and a DNA polymerase with proofreading activity to produce a DNA fragment which is complementary to the target and is at least 2 kb long. The present sequence encodes the spider silk protein from *Nephila clavipes*. The DNA fragment can be used to make fibres, films, woven articles, e.g. for use in parachutes, sails, body armour, and absorbers (e.g. of heavy metals, biological weapons, DNA, chemicals, flavours and fragrances). The high molecular weight (90-250 kD) of spider silk proteins can be produced on a commercial scale (at over 2 g/l cell mass). It has better tensile strength and elasticity than silkworm silk. Inclusion of both repetitive and non-repetitive regions ensures isolation of stable clones.

SQ Sequence 646 AA;

Query Match	85.98;	Score 2319;	DB 18;	Length 646;
Best Local Similarity	85.78;	Pred. No. 2e-152;		
Matches 484; Conservative	1;	Mismatches 14;	Indels 66;	Gaps 10;

OY	6	GGCGMGRRT-----RGYGLGCGAGCAQCGAGAAAAAAGAGCGCGYGTGCGAGGAGCGG	60
Db	60	ggggagagaaaaaaagagagggg ygggagagagag-aaaaaaagagggg ygg ygsqagtrggg	118
OY	61	AGAAAAAAG-----GAGCGGTGGLGSCGAGRGLGCGCGAG--AAAAAGVGCGGL	108
Db	119	agaaaaagagagggg ygggagagggg ygsqagtrgg ygggagagaaaaaagagaggg	178
OY	109	GGCGAGCGAGAAAAAAGGAGCGGTGGLGCGAGCGAGGSGCGAGAAAAAGCGCGTGG	168
Db	179	gggggagagaaaaaagagaggg ygg ygsqagtrgsggggagagaaaaaagagaggg ygl	238
OY	169	GSQCGAGRCGGLGCGCGAGAAAAAAGAGCGGTGGLGCGGAGCGGTGGLGSGGAGGGLGCG	228
Db	239	gsgggagtrgg ygggagagaaaaaagagaggg ygg yggagaggg ygg ygsqagtrgg ygg	288
OY	229	GAG--AAAAAGAGCGGGLG----CGAGAAAAAAGAGCGGTGGLGSGGAGRCGCGAGAA	282
Db	239	gagaaaaaaaagagggg ygggagagagagaaaaaagagggg ygg yggagtrgggagagaa	358
OY	283	AAAAAGAGCGGTGCGCGAGCGGTGGLGSGGAGRCGLGCGAGAAAAAAGAGAGC-----	356
Db	359	aaaaag-----gagaggg ygg ygsqagtrgg yggagag--aaaaaagagaggg ygg	406
OY	337	-----GGLGCGGAGAAAAAAGAGCGGTGGLGCGAGCGAGAAAAAAG	376
Db	407	ggggagggg ygg ygsqagtrgg ygggagagaaaaaagagggg ygggagagagag-----aaa	466
OY	377	AAAGCVRGGGTGGLGSCGAGRCGCGAGAAAAAAGAGCGGTGGLGCGGAGAGGLGCGAG	436
Db	463	aaagagaggg ygg ygsqagtrgsgggagagaaaaaagagaggg ygg ygggvgtrgg ygggag	522
OY	437	AAAAAGACGGGTGCGAGSASAAASAAASRLSSPQASRRYSASVNLVAGSPFNAAALSTI	496
Db	523	aaaagagggg ygg yggagaaaaaaaasrlsspgqasrrysavsnlvaspgfnasaa sstl	582
OY	497	SNVVOIGASNPGLSGCDVLITALL	521
Db	583	snvvsq qasnp gscdvl qall	607

RESULT	2
AAW53346	
ID	AAW53346 standard; Protein; 718 AA

XX AAW53346;
AC

DT 06-JUL-1998 (first entry)

DE Nephila clavipes spider silk protein.

kw Spider; *Nephila clavipes*; silk protein; tandem repeat; fibre; dragline,

XX

XX

XX

XX

XX

PR 20-APR-1990; 90US-0511792.

PR 19-APR-1995; 95US-0425069.

PA (UYWY-) UNIV WYOMING.

PI Hlman MB, Lewis RV, Xu M;

DR WPI; 1998-270437/24.

XX
XX

XX 01-1-201 20-24: 69ms

[illegible]

CC Invention. Spider silk prote

CC mixed and made into fibres

The present sequence represents a spider silk protein from the present invention. Spider silk proteins' and peptide fragments within the proteins, can be produced and purified independently and then be mixed and made into fibres that have higher tensile strengths and elasticity than naturally occurring fibres. The fibres can be used in mixed composites. The invention allows the two naturally occurring Nephlia clavipes silk proteins to be produced independently so that they can later be combined to form silk fibres of high tensile strength and elasticity.

SQ Sequence 718 AA;

Query Match	81.3%;	Score 2196;	DB 19;	Length 718;
Best Local Similarity	70.38%;	Pred. No. 6.4e-144;		
Matches 479; Conservative	4;	Mismatches 10;	Indels 188;	Gaps 15

[illegible]

	Query Match	81.3%	Score 2196:	DB 21:	Length 718:	
	Best Local Similarity	70.3%:	Pred. No. 6.4e-144:			
	Matches 479:	Conservative 4:	Mismatches 10:	Indels 188:	Gaps 15:	
OY	6 GGOQCRIRIRINRYGCLGCGAGCAGCAAAAAAAGAGAGCGYGCLGSQCAGRCGCAGAAA	65				
	:					
Db	22 gqgagagq---ggysglsgygaagagagaataaaagagaggyygylsgqagatrggqagaaa	78				
OY	66 AAAAGAGCGGCTGCLSGCAGRCGLCGCQAC--AAAAAGCVGGGSLG--- <td>119</td> <td></td> <td></td> <td></td> <td></td>	119				
Db	79 aaagagaggyygylsgqagatrgylsgqagaaaaaagagaggyygylgnqagatrgyqga	138				
OY	120 AAAAGAGACGCCGTGCLSGCAGRCSSCGCAGAAAAAAAG-----GAQGCGTG	166				
Db	139 aaaaagagaggyygylsgysagatrgylggysagaaaaaagagaggyygylggysagaggyg	198				
OY	167 GLGSGCARGRCGLCGCGAGA-----	185				
Db	199 glsgsagatrgylggysagaaaaaagagaggyglsgqagagagagagaggyygyl	258				
OY	186 -----AAAAAAGAGCGGCTGCGAGCGGCTGCGAGRCGLCGCQACA	232				
Db	259 gsqgagatrgysagaaaaaagagaggyygylggysagaggyygylsgqagatrgylggysag-	317				
OY	233 AAAAGAGACGCLG-----GQAGCAAAAAAGAGACGCCGTGCLGSQAGR---GCGACA-AA	284				
Db	318 aaaagagaggyglsgysagagagaaaaaagagaggyygylsgysagatrgylggysagavaaa	377				
OY	285 AAGGAGCGGCGG---GAGAGC-----GYGLGSCGAGRCGLCGCQACA	324				
Db	378 aagagaggyygylsgysagatrgysagaaaaaagagagatrgylgnqgagatrgylggysaga	437				
OY	335 AA-----AAAAAGAGAQ-----	336				
Db	438 aaaaaagagaggyygylgnqgagatrgysagaaaaaagagaggyygylsgysagatrgysagaaa	497				
OY	337 -----GCLGCGAGAAAAAAGAGCGGCLGCGAG	365				
Db	498 aaavagagatrgysagaggyygylsgysagatrgylggysagaaaaaagagaggyglggysag	557				
OY	366 OGAGAAAAAAGAGVROGCTGCLSGCAGRCGCAGAAAAAGAGCGCTGCLGCGCV	425				
Db	558 gqag-----aaaaagvtrgyygylsgysagatrgysagaaaaaagagaggyygylggysav	613				
OY	426 GAGGLGCGAGAAAAVGAOGCGCTGCGCASAAARLSSPQASSRVSAVSMLVASG	485				
Db	614 grgyglsgysagaaaaagaggyygylsgysagaaaaaarlsppqsarvsasvmnlvaeg	673				
OY	486 PTNSAALSTTISNVVQTIGAS	506				
Db	674 plnsaalstslsnvvsqlgas	694				
	RESULT	4				
	AARI4308					
xx	ID	AARI4308 standard; Protein: 718 AA.				
xx	AC	AARI4308;				
xx	DT	15-JAN-1992 (first entry)				
xx	DE	N.clavipes dragline silk protein-1.				
xx	KW	protein superfibre; major ampullate silk; orb web spider.				
xx	OS	Nephilila clavipes.				
xx	PN	EPA52925-A.				
xx	PD	23-OCT-1991.				
xx	PF	18-APR-1991; 91EP-0106217.				
xx						

[illegible]

	RESULT	6
CC	AAR9055	ID AAR9055 standard; Protein; 606 AA.
XX		
AC	AAR9055;	
XX		
DT	17-JAN-1997	(first entry)
XX		
DE	Spider dragline variant, DP-1B.9 polymer.	
XX		
KM	Spider; dragline protein; variant; monomer; polymer;	
KW	fibre forming region; Spidroin 1; Nephila clavipes; Dpl;	
KX	DP-1A analogue; fibre; high tensile strength; elasticity; clothing;	
XX	rope; surgical suture; implant; reinforcement; film; coating.	
XX	Synthetic.	
OS		
PN	MO9429450-A2.	
XX		
PD	22-DEC-1994.	
XX		
PF	15-JUN-1994:	94WO-US06689.
XX		
PR	15-JUN-1993:	93US-0077600.
XX		
PA	(DUPO) DU PONT DE MEMOURS & CO E I.	
P1	Fahnestock SR;	
XX		
DR	WPJ: 1995-036479/05.	
PT	New synthetic variants of spider dragline protein - for making	
PT	fibres useful as clothing, surgical silk, plastic reinforcement	
PT	etc., also related DNA, vectors and transformed cells	
XX		
PS	Disclosure; Page 88-90; 168pp; English.	
CC	This sequence represents a synthetic spider dragline variant polymer,	
CC	DP-1B.9. The sequence of the DP-1B.9 monomer is given in AAR9054.	
CC	The polypeptide monomer is a variant based on a consensus sequence	
CC	derived from the fibre forming regions of spider dragline protein,	

CC esp. the natural protein 1 (Spidroin 1) from *Nephila clavipes*. The
CC DNA sequence encoding the monomer may be used in the recombinant
CC production of the variant protein in a recombinant host, e.g. *E. coli*
CC or *Bacillus subtilis*. Synthetic analogues of DPI were designed to mimic
CC the repeating consensus sequence of the natural protein and the pattern
CC of variation among individual repeats. This monomer exhibits all of the
CC regularities of (1)-(5) below. In addition, it exhibits a regularity o
CC the natural sequence which is not shared by DP-1A, namely that a repeat
CC in which both GYG and GRG are deleted is generally preceded by a repeat
CC lacking the entire poly-alanine repeat, with one intervening repeat.
CC The sequence of DP-1B matches the natural sequence more closely over
CC a more extended segment than does DP-1A. The individual repeats differ
CC from the consensus sequence given in AAM06201 according to the pattern:
CC (1) the poly-alanine sequence varies in length from 0-7
CC residues; (2) when the entire poly-alanine sequence is deleted,
CC so also is the surrounding sequence encompassing AGRGGIGGGAGAGG;
CC encompass integral multiples of three consecutive residues;
CC (4) deletion of GYG is generally accompanied by deletion of GRG
CC in the same sequence; and
CC (5) a repeat in which the entire poly-alanine sequence is
CC deleted is generally preceded by a repeat containing six alanine
CC residues.
CC The proteins may be used to produce fibres of high tensile strength and
CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials
CC for implants, plastic reinforcements, films, coatings, etc.

Query Match	69.8%	Score 1883.5	DB 16	Length 606
Best Local Similarity	69.6%	Pred. No. 1.7e-122		
Matches 412	Conservative 3	Mismatches 26	Indels 151	Gaps 16
QY 6 GGOOMGRIRI-----RCGCGTGGCAGCAGCAAAAAAGCAAGCGGTGGTGGCAGCGCG 60				
Db 21 gqgqgaaaaaaagaaagaaqgqgglsgsqgaqgaq aaaaaaagaaqggygglsgsqgaqrsgg 79				
QY 61 AGAAAAAG-----GACGGGTGTLGSGCAGRGTGGCAGC--AAAAAGTCQ 105				
Db 80 agaaaaaagaaqaaqggygglsgsqgaqggygglsgsqgaqrgrglsgsqgaqaaaaaagaaq 139				
QY 106 GGLGGCGAGCGCGAGAAAAAGCAGCGCGTGGTGGCAGRGSAGCGAGAAAAAG----- 158				
Db 140 gglsgsqgaqgaaagaaaaaagaaqggygglsgsqgaqr-----gqgaaqaaaaaagaaqggy 196				
QY 159 -----GAGCGGTGTLGSGCAGRGTGGCAGCA----- 185				
Db 157 gglsgsqgaqggygglsgsqgaqrgrglsgsqgaqaaaaaagaaqgqgglsgsqgaqaa 256				
QY 186 -----AAAAAGCAGCGGTGTLGGCGAGCGGTGTLGSGQ 218				
Db 257 aaaqgaqggygglsgsqgaqrgrggaagaaaaaagaaqggygglsgsqgaqgygglsgq 316				
QY 219 GAGRGTGGLGCGAG-----AAAAAGCAGCGTGGC-----GCGAGAAAAAGCAGCGGTGTLGSGCA 272				
Db 317 gaqrgrglsgsqgaagaaaaaagaaqggygglsgsqgaqggaagaaagaaqggygglsgsqga 376				
QY 273 GRCGCGAGAAAAAGCAGCGGTG-----GCGCGGTGTLGSGCAGRGTGGCAGAAAAAA 329				
Db 377 grgqggaagaaaaaagaaqggygglsgsqgaqggygglsgsqgaqrgrglsgsqgaq aaaaa 435				
QY 330 AAGGAGCGCGC-----GCGAGAAAAAGCAGCGTGGC-----GCGAGCAAAAAAGAAAGV 382				
Db 436 aaqgaqggyglsgsqgaqgaagaaaaaagaaqggygglsgsqgaqrgrggaagaaaaaagaa 494				
QY 383 RCGGTGTLGSGCAGR-----GCG-----GCG 401				
Db 485 gqggygglsgsqgaqggygglsgsqgaqrgrglsgsqgaagaaaaaagaaqggygglsgsqgaq 554				
QY 402 AGAAAAAGCAGCGGTGTLGGCGVACGTGGCAGAAAAAGCAGCGTGGC 453				
Db 555 agaaaaaagaaqggygglsgsqgaqrgrgga aaaaaaagaaqggygglsgsqga 605				

```

RESULT 7
AAV40102
ID AAV40102 standard: protein; 606 AA.
XX
AC AAV40102;
XX
DT 19-NOV-1999 (first entry)
XX
DE Polymer of an analogue of spider silk protein spidroine major 1.
XX
KM Spider silk protein; spidroine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
KM hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephila clavipes.
XX
FH Key Location/Qualifiers
FT Peptide 1..101
FT /note= "monomer unit"
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PF 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA ) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudeau JP;
XX
DR WPI: 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PS for hair or skin care, in make-up or sunscreens
PS Claim 8; Fig 6B; 32pp; French.
XX
XX The present sequence represents a polymer of an analogue of the spider
CC silk protein spidroine major 1. The protein improves the moisturizing/
CC softening action of the compositions. The protein, and its fragments
CC are used in cosmetic or dermatological compositions. These compositions
CC have use as hair or skin care products; and make-up or sunscreens.
CC As the protein is a good, persistent film-formers on the skin
CC of low surface density, it can be used for delivery of active
CC agents that are generally difficult to administer, e.g. vitamins,
CC hormones, moisturizers or agents for treating disorders of the
CC skin and hair.
XX
SQ Sequence 606 AA:

```

```

Query Match 69.8%; Score 1883.5; DB 20; Length 606;
Best Local Similarity 69.6%; Pred. No. 1.7e-122;
Matches 412; Conservative 3; Mismatches 26; Indels 151; Gaps 16;

```

```

OY 6 GCGQMKRIRI-----RCVCGLCGCGAGCGAGAAAAAGAGCGGTGCTGSGAGRGCGG 60
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 21 ggggsggaaaaaaggagggg1gsggagggag-aaaaaaggagggg1gsggagrg99g 79
OY 61 AGAAAAAG-----GAGGCGYGTGSGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 105
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 80 agaaaaaaggagggg1gsggagggg1gsggaggg1gsggaggg1gsggaggg1gsggag 139
OY 106 GCGGCGAGCGAGAAAAAGAGCGGTGCTGSGAGRGCGGCGAGAAAAAG----- 158
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 140 ggl1gsggaggggaaaaaaggagggg1gsggaggr-----gggaggaaaaaaggagggg 196
OY 159 -----GAGGCGYGTGSGGAGRGGLGCGGAG----- 185

```

```

DB 197 ggl1gsggagggg1gsggaggg1gsggaggaaaaaaaggaggg1gsggagggagaa 256
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 186 -----AAAAAGAGCGGTGCTGCGAGCGGTGCTGCG 218
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 257 aaggagggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg 316
OY 219 GAGRGGLGCGAG-AAAAAGAGCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 272
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 317 ggg1gsggaggaaaaaaaggaggg1gsggaggg1gsggaggg1gsggaggg1gsggag 376
OY 273 GCGGCGAGCGAGAAAAAGAGCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 329
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 377 ggg1gsggaggaaaaaaaggaggg1gsggaggg1gsggaggg1gsggaggg1gsggag 435
OY 330 AAGGAGCGGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 382
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 436 aaggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggag 494
OY 383 RCGGTGCTGCGAGRG-----GCG 401
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 495 ggg1gsgg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggag 554
OY 402 AGAAAAAGAGCGGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 453
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 555 agaaaaaaggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg1gsggaggg 605

```

```

RESULT 8
AAV40101
ID AAV40101 standard: protein; 606 AA.
XX
AC AAV40101;
XX
DT 19-NOV-1999 (first entry)
XX
DE Polymer of an analogue of spider silk protein spidroine major 1.
XX
KM Spider silk protein; spidroine major 1; cosmetic; make-up;
KW dermatological compositions; hair care; skin care; sunscreen;
KM hormone; moisturizer; skin disorder; skin disorder.
XX
OS Synthetic.
OS Nephila clavipes.
XX
FH Key Location/Qualifiers
FT Peptide 1..101
FT /note= "monomer unit"
XX
PN FR2774588-A1.
XX
PD 13-AUG-1999.
XX
PF 11-FEB-1998; 98FR-0001614.
XX
PR 11-FEB-1998; 98FR-0001614.
XX
PA (OREA ) L'OREAL SA.
XX
PI Philippe M, Garson JC, Arraudeau JP;
XX
DR WPI: 1999-510729/43.
XX
PT Cosmetic or dermatological composition containing spider silk protein,
PS for hair or skin care, in make-up or sunscreens
PS Claim 8; Fig 5B; 32pp; French.
XX
XX The present sequence represents a polymer of an analogue of the spider
CC silk protein spidroine major 1. The protein improves the moisturizing/
CC softening action of the compositions. The protein, and its fragments
CC are used in cosmetic or dermatological compositions. These compositions
CC have use as hair or skin care products; and make-up or sunscreens.

```

As the protein is a good, persistent film-formers on the skin of low surface density, it can be used for delivery of active agents that are generally difficult to administer, e.g. vitamins, hormones, moisturizers or agents for treating disorders of the skin and hair.

Sequence 606 AA:

Query Match 69.6%; Score 1879.5; DB 20; Length 606;
Best Local Similarity 69.5%; Pred. No. 3.3e-122;
Matches 411; Conservative 3; Mismatches 26; Indels 151; Gaps 16;

```

QY 6 GCGQMGRIIRI-----RCYGLGCGAGCGAGAAAAAGAGCGCGTGLGSGCGRCGCG 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 23 gggagagaaaaaagagggg1gsgagggag-aaaaaagagggg1gsgagggg 81
QY 61 AGAAAAAG-----GAGCGTGLGSGAGRCGLGCGAG--AAAAAGVGQ 105
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 82 agaaaaaagagggg1gsgagggg1gsgaggg1gsgaggg1gsgaggg1gsgag 141
QY 106 GGLGCGAGCGAGAAAAAGAGCGCGTGLGSGCGRCGCGAGAAAAAG----- 158
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 142 gglgsgagggagaaaaaagggg1gsgaggr---gggagagaaaaaagggg 198
QY 159 -----GAGCGTGLGSGAGRCGLGCGAGA----- 185
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 gglgsgagggg1gsgaggr1gsgagagaaaaaagaggg1gsgagggagaa 258
QY 186 -----GAGCGTGLGSGAGRCGLGCGAGA----- 218
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 259 aaagagggg1gsgaggr1gsgagagaaaaaagaggg1gsgagggg1gsgag 318
QY 219 GAGRCGLGCGAG--AAAAAGAGCGCGTGLGSGCGRCGLGCGAGAAAAAG----- 272
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 319 gggg1gsgagggagaaaaaaggg1gsgagggagaaaaaagggg1gsgag 378
QY 273 GRCGCGAGAAAAAGAGCGCGTGLGSGCGRCGLGCGAGAAAAAG----- 329
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 379 g1gsgagggagaaaaaaggg1gsgagggg1gsgaggg1gsgaggg1gsgag 437
QY 330 AAGGAGCGGCG-----GCGAGAAAAAGAGCGCGTGLGSGCGRCGLGCGAGAAAAAG 382
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 438 aaagaggg1gsgagggagaaaaaagaggg1gsgaggr1gsgagaaagaa 496
QY 383 RGGCGTGLGSGAGRG-----GCGAG----- 401
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 497 gggg1gsgagggg1gsgaggr1gsgaggg1gsgaggg1gsgaggg1gsgag 556
QY 402 AGAAAAAGAGCGCGTGLGCGCGTGLGCGAGAAAAAGAGCGCGTGLGCGAG 452
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 557 agaaaaaagaggg1gsgaggr1gsgaggr1gsgaggr1gsgaggr1gsgag 606

```

RESULT 9

AAR9057 standard: Protein: 604 AA.

AAR9057:

17-JUN-1997 (first entry)

Spider dragline variant, DP-1B.16 polymer.

Spider dragline protein; variant; monomer; polymer;

fibres forming region; Spidroin 1; Nephila clavipes; Dp1; mimic;

DP-1A analogue; fibre; high tensile strength; elasticity; clotting;

rope; surgical suture; implant; reinforcement; film; coating.

Synthetic.

MO9429450-A2.

PD 22-DEC-1994.

XX 15-JUN-1994; 94MO-US06689.

XX 15-JUN-1993; 93US-0077600.

PA (DUPO) DU PONT DE NEMOURS & CO E. I.

PI Fahnstock SR;

XX WPI; 1995-036479/05.

PT New synthetic variants of spider dragline protein - for making

PT fibres useful as clothing, surgical silk, plastic reinforcement

PS etc., also related DNA, vectors and transformed cells

XX Disclosure; Page 106-108; 168pp; English.

CC This sequence represents a synthetic spider dragline variant polymer,

CC DP-1B.16. The sequence of the DP-1B.16 monomer is given in AAR9056.

CC The polypeptide monomer is a variant based on a consensus sequence

CC derived from the fibre forming regions of spider dragline protein,

CC esp. the natural protein 1 (Spidroin 1) from Nephila clavipes. The

CC DNA sequence encoding the monomer may be used in the recombinant

CC production of the variant protein in a recombinant host, e.g. E. coli

CC or Bacillus subtilis. Synthetic analogues of DP1 were designed to mimic

CC the repeating consensus sequence of the natural protein and the pattern

CC of variation among individual repeats. This monomer exhibits all of the

CC regularities of (1)-(5) below. In addition, it exhibits a regularity of

CC the natural sequence which is not shared by DP-1A, namely that a repeat

CC in which both GYG and GRG are deleted is generally preceded by a repeat

CC lacking the entire poly-alanine repeat, with one intervening repeat.

CC The sequence of DP-1B matches the natural sequence more closely over

CC a more extended segment than does DP-1A. The individual repeats differ

CC from the consensus sequence given in AAM06201 according to the pattern:

CC (1) the poly-alanine sequence varies in length from 0-7

CC residues; (2) when the entire poly-alanine sequence is deleted,

CC so also is the surrounding sequence encompassing AGRGGLGCGAGAGC;

CC (3) aside from the poly-alanine sequence, deletions usually

CC encompass integral multiples of three consecutive residues;

CC (4) deletion of GYG is generally accompanied by deletion of GRG

CC in the same sequence; and

CC (5) a repeat in which the entire poly-alanine sequence is

CC deleted is generally preceded by a repeat containing six alanine

CC residues.

CC The proteins may be used to produce fibres of high tensile strength and

CC elasticity, suitable for clothing, rope, surgical sutures, biomaterials

CC for implants, plastic reinforcements, films, coatings, etc.

Sequence 604 AA:

Query Match 69.3%; Score 1871.5; DB 16; Length 604;
Best Local Similarity 69.8%; Pred. No. 1.2e-121;
Matches 411; Conservative 3; Mismatches 26; Indels 149; Gaps 17;

```

QY 6 GCGQMGRIIRI-----RCYGLGCGAGCGAGAAAAAGAGCGCGTGLGSGCGRCGCG 60
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 23 gggagagaaaaaagaggg1gsgagggag-aaaaaagaggg1gsgagggg 81
QY 61 AGAAAAAG-----GAGCGTGLGSGAGRCGLGCGAG--AAAAAGVGQ 105
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 82 agaaaaaagaggg1gsgaggg1gsgaggg1gsgaggg1gsgaggg1gsgag 141
QY 106 GGLGCGAGCGAGAAAAAGAGCGCGTGLGSGCGRCGCGAGAAAAAG----- 158
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 142 gglgsgagggagaaaaaaggg1gsgaggr---gggagagaaaaaagggg 198
QY 159 -----GAGCGTGLGSGAGRCGLGCGAGA----- 188
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 199 gglgsgagggg1gsgaggr1gsgagagaaaaaagaggg1gsgagggagaa 258
QY 189 AAGGAGCGGCGTGLG-----GCGAGCGCGTGLGSGOGA 220

```


DE	Polymer of an analogue of spider silk protein spidroline major 1.
XX	
KW	Spider silk protein; spidroline major 1; cosmetic; make-up;
KM	dermatological compositions; hair care; skin care; sunscreen;
KV	hormone; moisturizer; skin disorder; skin disorder.
XX	
OS	Synthetic.
OS	Nephila clavipes.
XX	
PH	Key
FT	Location/Qualifiers
FT	1..101
FT	/note= "monomer unit"
XX	
PN	FR2774588-A1.
XX	
PD	13-AUG-1999.
XX	
PF	11-FEB-1998; 98FR-0001614.
XX	
PR	11-FEB-1998; 98FR-0001614.
XX	
PA	(OREA) L'OREAL SA.
XX	
PI	Phillippe M, Garson JC, Arraudau JP.
XX	
DR	WPI, 1999-510729/43.
XX	
PT	Cosmetic or dermatological composition containing spider silk protein,
PT	for hair or skin care, in make-up or sunscreens
XX	
XX	Claim 8; Flg 4B; 32pp; French.
XX	
PS	The present sequence represents a polymer of an analogue of the spider
CC	silk protein spidroline major 1. The protein improves the moisturizing/
CC	softening action of the compositions. The protein, and its fragments
CC	are used in cosmetic or dermatological compositions. These compositions
CC	have used as hair or skin care products, and make-up or sunscreens.
CC	As the protein is a good, persistent film-formers on the skin
CC	of low surface density, it can be used for delivery of active
CC	agents that are generally difficult to administer, e.g. vitamins,
CC	hormones, moisturizers or agents for treating disorders of the
CC	skin and hair.
XX	
XX	
SQ	Sequence 606 AA:
	Query Match 68.8%; Score 1558.5; DB 20; Length 606;
	Best Local Similarity 73.8%; Pred. No. 9, 2e-121;
	Matches 409; Conservative 5; Mismatches 31; Indels 109; Gaps 18;
QY	6 GGGOMGRINI-----RGVGLGGGAGGAGCAAAAAAG-----GAGGGGCG 48
DB	38 gggaggaataaaaggaaggaggggagggaggaataaaaggaagggygggaggaaggaggg 97
QY	49 LGSQGAGRGCGAGAAAAAGAGAGGCGYGLGSQGAGRGGLGGGCGAG--AAAAAGGCGG 106
DB	98 lgsqgagrgggaggaataaaaggaagggygggaggaagggygggaggaataaaaggaagg 157
QY	107 GLGCGAGGAGCAAAAAAGAGGAGCGGTGGLGSQGAGRG-----GGGCGAGCAAAA 156
DB	158 gltgsqgagrggaataaaaggaagggygggaggaagggygggaggaagggyggagaaaaa 217
QY	157 AGGAGCGGCGGAGGAGGAGGGLGCGGAGCA-----AAAAAG 192
DB	218 aggaaggaggygggaggaagggygggaggaataaaaggaagggyggaggaaggagaaaaa 277
QY	193 GAGGCGGGLGSGAGCGGCGYGLGSQGAGRGGLGCGGAGCAAAAAG-----G 238
DB	278 gaggaggygggaggaagggygggaggaagggygggaggaagggygggaggaagggyggg 334
QY	239 AGCGGAGCGGAG--AAAAAGGAGGCGGCGGGLGSGAGRGGAGCAAAAAAGAGGCGGCG 296
DB	335 agtgggaggaataaaaggaagggygggaggaagggygggaggaagggygggaggaagggygg 388

```

QY 237 --QGHGCGGTCGTCGACGKCGTGLGCGACGAAAAAA--AAG-----GAGCGGTLGCG 343
    |||||||
Db 389 gsgqagqggygylgsqgaqr---gqgqagaaaaaagqggygylgsqgaqrqglgsq 445
    |||||||

QY 344 AG-AAAAAAGCAGCGGCGGCGAGCGAAGAAAAAA-----AGCVRGCGTCGTCG 393
    |||||||
Db 446 agaaaaaaggaagqgylgsqgaqgaagaaaaaagqggygylgsqgaagqggyglgsq 505
    |||||||

QY 394 GAGRGGCGAGAAAAAAGCAGAGCGTCGTCGTCGAGAGGLGCGGAGAAAAA--GAGCGGCG- 449
    |||||||
Db 506 gqgrygqgagaaaaaagqggygylgsqgaqrqglgsqgaagaaaaaaggaagqgylgs 565
    |||||||

QY 450 -GVGSGASASASAAA 462
    || || || || ||
Db 566 qgagqgagaaaaaa 579

RESULT 12
AAR80168
ID AAR80168 standard; Protein; 831 AA.
XX
XX AAR80168;
XX AC
XX DT 01-MAY-1996 (first entry)
XX
XX pMIS1 M1SP spider silk protein insert product.
XX
XX Spider silk; repeat unit; consensus; minor ampullate silk protein;
KW spideroin; M1SP; orb web spider; dragline.
XX
XX Nephila clavipes.
XX
XX
XX Key Location/Qualifiers
FH Misc-difference 272..285
FT /note- "represented as indeterminate, corresponds
FT to a highly compressed GC rich region in
FT AAG98470 which could not be sequenced"
XX
XX W09525165-A1.
XX
XX 21-SEP-1995.
XX
XX 14-MAR-1995; 95MO-US03139.
XX
XX 14-MAR-1994; 94US-0209747.
XX
XX (UYWY-) UNIV WYOMING.
XX
XX Colgin M, Lewis RV;
XX
XX WPI; 1995-336970/43.
XX
XX N-PSDB; AAO98470.
XX
XX Polypeptide(s) comprising repeated unit amino acid sequences, also
XX PD cDNAs - derived from minor ampullate spider silk proteins and used
XX PT to form spider silk fibres
XX
XX Claim 2; Fig 1; 86pp; English.
XX
XX
XX This sequence represents the product of cDNA clone, pMIS1, encoding
XX the orb web spider minor ampullate silk protein M1SP1 has been
XX identified and sequenced. Repeat unit peptides of M1SP1 may form
XX part of a larger polypeptide with an amino terminus (see AAR80184-85)
XX and a carboxy terminus (see AAR80186-89). Repeat unit peptides make up
XX spider silk proteins (spidroins) which in turn aggregate to form the
XX silk fibres. Spider silk fibres have high tensile strength and
XX significant elasticity. An isolated cDNA clone of a silk protein
XX encoding sequence is of use to produce the protein at high yields
XX using recombinant DNA technology.
XX
XX Sequence 831 AA;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 14:56:41 ; Search time 48.42 Seconds

(Without alignments)
219,669 Million cell updates/sec

Title: US-09-490-291-8

Perfect score: 2700
Sequence: 1 MASMTGGCGMGRIRIRGCG.....GLSCDVLIALGLHHHHH 528Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2273	84.2	747	4	US-09-034-177-3 Sequence 3, Appl1
2	2196	81.3	718	1	US-08-425-069-2 Sequence 2, Appl1
3	2196	81.3	718	2	US-08-317-844B-2 Sequence 2, Appl1
4	1343.5	49.8	832	1	US-08-209-747-2 Sequence 2, Appl1
5	1343.5	49.8	832	1	US-08-458-298-2 Sequence 2, Appl1
6	1144	42.4	595	1	US-08-425-069-4 Sequence 4, Appl1
7	1144	42.4	595	2	US-08-317-844B-4 Sequence 4, Appl1
8	972	36.0	1177	1	US-07-609-716-31 Sequence 31, Appl1
9	972	36.0	1177	1	US-08-175-155-29 Sequence 29, Appl1
10	972	36.0	1177	1	US-08-477-509B-64 Sequence 64, Appl1
11	972	36.0	1177	2	US-08-707-237A-35 Sequence 35, Appl1
12	972	36.0	1177	3	US-08-482-085B-64 Sequence 64, Appl1
13	972	36.0	1177	4	US-08-475-411A-31 Sequence 31, Appl1
14	972	36.0	1177	4	US-08-478-029A-31 Sequence 31, Appl1
15	962.5	35.6	1059	1	US-08-175-155-48 Sequence 48, Appl1
16	962.5	35.6	1059	2	US-08-707-237A-54 Sequence 54, Appl1
17	962.5	35.6	1101	1	US-08-477-509B-83 Sequence 83, Appl1
18	962.5	35.6	1101	3	US-08-482-085B-83 Sequence 83, Appl1
19	929	34.4	1332	1	US-07-609-716-41 Sequence 41, Appl1
20	929	34.4	1332	4	US-08-475-411A-41 Sequence 41, Appl1
21	929	34.4	1332	4	US-08-478-029A-41 Sequence 41, Appl1
22	922.5	34.2	1038	1	US-07-609-716-36 Sequence 36, Appl1
23	922.5	34.2	1038	4	US-08-475-411A-36 Sequence 36, Appl1
24	922.5	34.2	1038	4	US-08-478-029A-36 Sequence 36, Appl1
25	921.5	34.1	766	1	US-08-175-155-53 Sequence 53, Appl1
26	921.5	34.1	766	1	US-08-477-509B-88 Sequence 88, Appl1
27	921.5	34.1	766	2	US-08-707-237A-61 Sequence 61, Appl1

28	921.5	34.1	766	3	US-08-482-085B-88 Sequence 88, Appl1
29	921.5	34.1	979	1	US-08-477-509B-89 Sequence 89, Appl1
30	921.5	34.1	979	3	US-08-482-085B-89 Sequence 89, Appl1
31	921.5	34.1	1050	1	US-08-175-155-54 Sequence 54, Appl1
32	921	34.1	745	2	US-09-010-928B-28 Sequence 28, Appl1
33	921	34.1	870	2	US-09-010-928B-2 Sequence 2, Appl1
34	911.5	33.8	1018	1	US-08-089-862-11 Sequence 11, Appl1
35	911.5	33.8	1018	1	US-08-587-333-18 Sequence 18, Appl1
36	911.5	33.8	1018	5	PCT-US94-07776-16 Sequence 16, Appl1
37	895.5	33.2	1011	1	US-08-477-509B-94 Sequence 94, Appl1
38	895.5	33.2	1011	3	US-08-482-085B-94 Sequence 94, Appl1
39	895.5	33.2	1170	1	US-08-175-155-59 Sequence 59, Appl1
40	895.5	33.2	1170	2	US-08-707-237A-66 Sequence 66, Appl1
41	890.5	33.0	738	3	US-08-864-038A-3 Sequence 3, Appl1
42	889.5	32.9	784	1	US-07-609-716-48 Sequence 48, Appl1
43	889.5	32.9	784	4	US-08-475-411A-48 Sequence 48, Appl1
44	889.5	32.9	784	4	US-08-478-029A-48 Sequence 48, Appl1
45	882	32.7	2100	1	US-08-477-509B-80 Sequence 80, Appl1

ALIGNMENTS

RESULT 1
US-09-034-177-3
Sequence 3, Application US/09034177
Patent No. 6127146
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN FIBROUS PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
City: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,177
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0486 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: GI 1174414
US-09-034-177-3
Query Match 84.2% Score 2273; DB 4; Length 747;

[illegible]

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,069
FILING DATE: 19-Apr-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-106P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 205-8000
TELEFAX: (703) 205-8050
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 amino acids
type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-425-069-2

```

[illegible]

TISSUE TYPE: minor amputate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 1..309
US-08-209-747-2

Query Match 49.8%; Score 1343.5; DB 1; Length 832;
Best Local Similarity 47.9%; Pred. No. 1e-91;
Matches 343; Conservative 19; Mismatches 143; Indels 211; Gaps 25;

QY 2 ASMTGCGQMGRRIRIRYGGGLGGGAGGCGGAGAAAAAGAGGCGG-----G 48
DB 93 AAGAGGCGRG---AGGCGGGGAGAGAGAAAAAGAGAGAGGCGGAGAGAGAGAAAG 149
QY 49 LGSQAGRGCGQ-----GAGAAAAAGAGAGCGGCGGGLG-----SOGAGRGGLGCG-- 92
DB 150 AGAGGAGGCGGCGGAGAGAGAAAAAGAGAGAGCGGCGAGAGAGAGAGAGAGAGGCGG 209
QY 93 -----GAGAAAAAGGCGGCGGAGAGAGAAAAAGAGAGCGGCGGGLG----- 136
DB 210 YGAGAGAGAAAAAGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 269
QY 137 -----SOGAGRGCGS--GQAGAAAAAGAGAGCGGCGGGLG----- 169
DB 270 AAXXXXXXXXXXXXXXGAGAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 329
QY 170 -----SOGAGRGGLGCGQ-----GAGAAAAAG----- 192
DB 330 AGAAGAGAGAGAGAGGCGGCGGAGAGAGAAAAAGAGAGAGCGGCGGAGAGAGAGAGAA 389
QY 193 GAGCGGCGGGLGCGAGCGGCGGGLGCGGAGRGGLGCGGAGAGAGAGAGAGAGAGAGAG 247
DB 390 GAGAGCGGCGGCGGAG 448
QY 248 -----GAGAAAAAG 281
DB 449 YGAGAGAGAGAAAAAG 508
QY 282 AAAAAAGAGAGCGG-----GQAGCGGCGGCGGGLGCGAG----- 313
DB 509 AAAAAAG 568
QY 314 -----RGGLGCGGAGAGAAAAAGAGAGCGGGLGCGQ-----GAGAAAAAGAGAGAG 357
DB 569 GAG 627
QY 358 GAGCGGAG 417
DB 628 GAGCGGCGGAG 682
QY 418 GGLGCGGAG 465
DB 683 -----GAGAGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 738
QY 466 SSPQASRRVSSAVSNLVSAGPTNSAALSTISNVVSQLASNPGLSGCVLIQALL 521
DB 739 SSAEASSRISSAASTLVSGGTLNTAALPSVLDLFAQVAGASSPVLRQRS-LIQVLL 793

RESULT 5
US-08-458-298-2
Sequence 2, Application US/08458298
Patent No. 5756677

GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Colgin, Mark
TITLE OF INVENTION: CDNAS Encoding Minor Amputate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747

CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,747
FILING DATE: 14-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: N. clavipes
TISSUE TYPE: minor amputate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 1..309
US-08-458-298-2

Query Match 49.8%; Score 1343.5; DB 1; Length 832;
Best Local Similarity 47.9%; Pred. No. 1e-91;
Matches 343; Conservative 19; Mismatches 143; Indels 211; Gaps 25;

QY 2 ASMTGCGQMGRRIRIRYGGGLGGGAGGCGGAGAAAAAGAGGCGG-----G 48
DB 93 AAGAGGCGRG---AGGCGGGGAGAGAGAAAAAGAGAGAGGCGGAGAGAGAGAGAAAG 149
QY 49 LGSQAGRGCGQ-----GAGAAAAAGAGAGCGGCGGGLG-----SOGAGRGGLGCG-- 92
DB 150 AGAGGAGGCGGCGGAG 209
QY 93 -----GAGAAAAAGGCGGCGGAGAGAGAGAAAAAGAGAGCGGCGGGLG----- 136
DB 210 YGAGAGAGAGAAAAAGAA 269
QY 137 -----SOGAGRGCGS--GQAGAAAAAGAGAGCGGCGGGLG----- 169
DB 270 AAXXXXXXXXXXXXXXGAGAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 329
QY 170 -----SOGAGRGGLGCGQ-----GAGAAAAAG----- 192
DB 330 AGAAGAGAGAGAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 389
QY 193 GAGCGGCGGGLGCGAGCGGCGGGLGCGGAGRGGLGCGGAGAGAGAGAGAGAGAGAGAG 247
DB 390 GAGAGCGGCGGCGGAG 448
QY 248 -----GAGAAAAAG 281
DB 449 YGAGAGAGAGAAAAAG 508

QY	282	AAAAAGAGGCGGCGG-----GQAGGCGGCGGCGGCGGAG-----	31.3
Db	509	AAAAAGAGGAGGAGTCGACAGAGCAAGAAAGAGAGAGAGGTCGACARAGAAAA	56.6
QY	314	-----RGLGCGCGAGAAAAAAGAGCGGLGCG-----GAGAAAAAAGAGCG	35.7
Db	569	CAGAGCAGAGTCGCGRAG--AAGAGAGAAAAAGAGAGCGCGCGGTCGAGAGCAAAAAAGAGCG	62.2
QY	358	GLGGGAGGCGGAGAAAAAAGAGGVRGGCGGAGCGAGAGCGGAGAAAAAGAGAGCGGT	41.7
Db	628	GAGGTCGACAGAGAAAAAGAGAGAGAGAGCGGTGCA-----GAGAAAAAGAGAGAGCGT	68.8
QY	418	GGGCGGCGGAGGCGG-----GAGAAAGAGCGGCGGCGGCA-----SAASAAARL	46.5
Db	683	-----GCGAGAGGCGGCGGCGAGAGAGAAAAAAGAGAGGCGGCAKEICMGRCRVTYASTSRLL	73.8
QY	466	SSPQASVSVASVNLVAVSGPTNSAALSTSTSNVYSOIGANSPELSCDVLQALL	52.1
Db	739	SSAASSTISSAAATVLSGGLNLTALPLSVTSDLPFAOVGASASPVTIRORS--LIQVLL	79.3

RESULT 6
 US-08-425-069-4
 : Sequence 4: Application US/08425069
 : Patent No. 5728810
 : GENERAL INFORMATION:
 : APPLICANT: Lewis, Randolph V.
 : APPLICANT: Xu, Ming
 : APPLICANT: Himman, Michael B.
 : TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
 : TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 : TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 : NUMBER OF SEQUENCES: 69
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Birch, Stewart, Kolasch & Birch
 : STREET: 301 No. 5728810th Washington Street
 : CITY: Falls Church
 : STATE: Virginia
 : COUNTRY: U.S.A..
 : ZIP: 22046
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/425,069
 : FILING DATE: 19-APR-1995
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Murphy Jr., Gerald M
 : REGISTRATION NUMBER: 28,977
 : REFERENCE/DOCKET NUMBER: 1447-106P
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (703) 205-8000
 : TELEFAX: (703) 205-8050
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 595 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-425-069-4

Query Match	42.4%	Score 1144	DB 1	Length 595
Best Local Similarity	51.1%	Pred. No. 43-77		
Matches 313	Conservative 25	Mismatches 139	Indels 136	Gaps 30
QY	17	GTC-----GLGCGAGC-----GAG-AAAAAAAAAGCAGCGTGC-----GLGCGCAGC--	56	
db	3	GTCGCGCGCGTGCPCGCGPSCGCGSAAAAAAAAAAGCGCGCGCGCGCGCGCGCGTGC	62	

OY	57	----	GCOCAGAAAAAGAGCGCGTGGTGSOGAGGGLG--G0----	-GAGAAAAAGV----	103	
Db	63	PG00GPGSPGSA	AAAAAGSG00GPGGTGPRK00GPGGTGG00GPGSPGSGSA	AAAAASAE	122	
OY	104	-G0GGLG0G0GAG0GAG	AAAAAAGGAG00GYG----	GLGS0GAGRGSGSG0G--AGAAAAA	157	
Db	123	SG00GPGGTGPGQ-----	QGCGYGGPG00GPGGYGPG00GPGSGSA	AAAAAA	171	
OY	158	GGAGC----	GGYG-----	GLGS0GAGRGGLG0GAGAAAAAAGGAG00GYTGGTGG0GAG0G	210	
Db	172	SGPG00GPGGTGPG00GPGGTGPG00GPGSGSA	AAAAAASGPG00GPGGYGPG00PG	231		
OY	211	GYGGLGS0GAGRGGLG0GAGAAAAAGAG00GLG0GAG-----	AAAAAAGAGCQ	261		
Db	232	GY-----	GPR00GLSGPBGSA	AAAAAAGPG00GPGGTGPR00GPGSPGSGSA	AAAAAAG	285
OY	262	GGYG----	GLGS0GAG--RG0GAG--AAAAAAGGAG00GYG0GAG00GYGGLGS0GAGRG	315		
Db	286	GGYGG00GPGGTGPG00GPGSGAGS	AAAAAAGPG00GPGGTGPG00GPGG----	YG000	342	
OY	316	GLG0GAGAAAAAAGGAG00GGLG0GAG-----	AAAAAAG----	GAG0GSG	358	
Db	343	GPGGTGPGPSASAAAAAGPG00GPGGTGPG00GPGSGSA	AAAAAAGPGGTGPG00G	402		
OY	359	LG0GAGC----	GAGAAAAAAGGVR0G0GGLGS0GAG-----	RG0GAG--AAAA	407	
Db	403	PGGTAPG00GPGSPGSGSA	AAAAAAG--PGGTGPR--G00GPGTAPG00GPGSPGSGSA	AAA	459	
OY	408	AAGGAG0G0GTG----	GLG0G0V-----	GAGGLG-----	G0GAGAAAAVGAG0G	447
Db	460	AAASAGPGGTGPA00GPGPGTASASASAGGTGTPA00GPGAGTGPESAV	AAASAGASAG	519		
OY	448	YGVGSGASASAAASRLSSPQASRVSSAVSNLVA	SGPTNSAALSTJTNVY	SGTQASN	507	
Db	520	YG----	PGSQASAAASRLASPSDCAVAAVSNLTVSSGPTSSAALSSV	ISNAVS0TQASN	575	
OY	508	PGLSGCDVLTAL	520			
Db	576	PGLSGCDVLTAL	588			

RESULT 7
 US-08-317-844B-4
 Sequence 4, Application US/08317844B
 Patent No. 5989894
 GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph V.
 APPLICANT: Xu, Ming
 APPLICANT: Himman, Michael B.
 TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
 TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 NUMBER OF SEQUENCES: 62
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Birch, Stewart, Kolasch & Birch
 STREET: 301 No. 5989894th Washington Street
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22046
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/317,844B
 FILING DATE: 04-OCT-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M
 REGISTRATION NUMBER: 28,977

OY 516 LIOALLGHHH 526
Db 1160 RYOLSGRHYH 1170

RESULT 9

US-08-175-155-29
; Sequence 29, Application US/08175155
; Patent No. 5641648
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W.
; APPLICANT: Dorman, Mary A.
; TITLE OF INVENTION: Methods for Preparing Synthetic
; TITLE OF INVENTION: Repetitive DNA
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,155
; FILING DATE: 29-DEC-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertam I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-5/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-175-155-29

Query Match 36.0%; Score 972; DB 1; Length 1177;
Best Local Similarity 44.1%; Pred. No. 3e-64;
Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;

OY 17 GYGLGCGGAGGAGAAAAAG-----GAGGCGYGLGCGGAGRG-GGAGGAAAAAG 70
Db 633 GAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 692
OY 71 AGGCGYGLG-OGAGRGCGGAGGAGAAAAAGYGGGCGGAGGAGAAAAAGAGG 129
Db 693 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 752
OY 130 GYGLGCGGAGGAGGAGAAAAAG-GAGGCGYGLG-OGAGRGCGGAGGAG 184
Db 753 GAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 812
OY 185 AAAAAAG---GAGGCGYGLG-GGAGGCGYGLG-OGAGRGCGGAGGAGAAAAAG- 237
Db 813 AGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 872
OY 238 ----GAGGCGGAGGAGAAAAAGGAGGCGYGLGSGAGRG-----GGAGAGAA 284
Db 873 GSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAG 932

OY 285 AAGAGGCGYCGGAGCGGCGYGLGSGGAGRGGLGCGGAGAAAAAGAGAGCGGAGG 344
Db 933 SGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGS 992
OY 345 GAAAAAGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 398
Db 993 GAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 1052
OY 399 ---GGAGAAAAAGAGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 455
Db 1053 AGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 1112
OY 456 SAASAAASRLSPQASSRVSAVSNLVAASGPTNSAALSTISINVSQIGASNBSGCDV 515
Db 1113 GAGSAG-----AGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGMDPG----- 1159
OY 516 LIOALLGHHH 526
Db 1160 RYOLSGRHYH 1170

RESULT 10

US-08-477-509B-64
; Sequence 64, Application US/08477509B
; Patent No. 5770697
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A
; APPLICANT: Cappello, Joseph
; APPLICANT: Crissman, John W
; APPLICANT: Dorman, Mary A
; TITLE OF INVENTION: No. 5770697el Peptides Comprising Repetitive
; TITLE OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
; NUMBER OF SEQUENCES: 112
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,509B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,155
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,049
; FILING DATE: 22-APR-1993
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-7/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1177 amino acids
; TYPE: amino acid


```

OY 399 ---GGGAGAAAAAAGGAAGGTGGTGGGGGVAGGLGGCGAGAAAANVGAOGGTGYGCSGA 455
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1053 AGAGSGAGAGCAGACGAGSAGACSCGAATGCAGAGSGACAGAGSGACGAGSGAGAGSGA 1112
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 456 SASASAAARSLLSPQASSRVSSAVSNLVAAGPTNSALSTISNVYSIGASNPGLSGCDV 515
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1113 GAGSGAG-----AGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAMDPG----- 1159
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 516 LIOALLCHHHH 526
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1160 RYOLAGRYHY 1170
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-08-482-085B-64
Sequence 64, Application US/08482085B
Patent No. 6018030
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Richardson, Charles
APPLICANT: Chambers, James
APPLICANT: Causey, Stuart
APPLICANT: Pollock, Thomas J.
APPLICANT: Cappello, Joseph
APPLICANT: Crissman, John W.
TITLE OF INVENTION: No. 6018030el Peptides Comprising Repetitive
NUMBER OF INVENTION: Units of Amino Acids and DNA Sequences Encoding the Same
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Honbach, Teat, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,085B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,049
FILING DATE: 22-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,155
FILING DATE: 29-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Treccarlin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-6/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 1177 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-482-085B-64

```

```

Query Match Similarity      36.0%; Score 972; DB 3; Length 1177;
Matches 243; Conservative 30; Mismatches 224; Indels 54; Gaps 17;
Pred. No. 3e-64;

QY 17 GYGGIGCGGAGGAGGAGAAAAAAG-----GAGGCGYGGIGLGSQGAGRG-CGGAGAAAAAAG 70
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 633 GAGAGSGAGAGSGAGAGSGCAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSG 692
QY 71 AGCGGCGGLGS-QGAGRGGLGCGGAGAAAAAAGAGVCGGGLGCGGAGCGGAGAAAAAAGCAQ 129
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 693 AGAGGGAAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGDA 752
QY 130 GCGYGLGS-----GAGRGSGCGGAGAAAAAAG-GAGCGGCGGLGS-QGAGRGGLGCGGAG 184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 753 GAGSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 812
QY 185 AAAAAAAG--GAGGCGYGGIGL-GGAGCGGCGYG-GLGS-QGAGRGGLGCGGAGAAAAAG- 237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 AGSGAGAGSGAGAGSGAGAGSGAGAGSGCAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 872
QY 238 ----GAGCGGLGCGGAGAAAAAAGAGCGCGYGGIGLGSQGAGRG-----CGGAGAAAA 284
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 873 GSGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 932
QY 285 AAGGAGCGCGGCGGAGCGCGYGGIGLGSQGAGRGGLGCGGAGAAAAAAGAGAGCGGLGCGGA 344
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 933 SGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 992
QY 345 GAAAAAAGAGCGGGLGCGGAGCGAGAGA-----AAAAAAGAGVCGGCGYGGIGLGS-QGAGRG 398
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 993 GAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSG 1052
QY 399 ---GGGAGAAAAAAGAGAGCGCGTGTGGLGCGCGVAGAGLGGCGGAGAAAAAVGAGCGGCGV 455
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1053 AGAGGAGAGSGAGAGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGAGSGAGA 1112
QY 456 SAASAASRLSSPQASSRVSASVMSLVASGPTNSAALSTISNVVSQLGASNPGLSGCDV 515
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1113 GAGSGAG-----AGSGAGAGSGAGAGSGAAGYGAGAGSGAGAGSGAGAGSGAGACAMPG----- 1159
QY 516 LIQALLGHHH 526
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1160 RYQLSAGRHY 1170

RESULT 13
US-08-475-411A-31
; Sequence 311, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```



```

Db      813  AGCGAGAGSSGAGAGSGAGAGSGAGAGCGAAATGCGAGAGSGAGAGSGAGAGSGAGAGSGAGA  872
Oy      238  ----GACGGGGLGCGGACAGAAAAAAGAGCGGGGGLGSGGAGRG-----GCGAGAAA  284
Db      873  GSGAGSGAGAGSGAGAGSGAGAGSGAGAGCGAAGTGCAGAGSGAGAGSGAGAGSGAGAGSGAGAG  932
Oy      285  AAGGAGCGGCGGCGGAGCGGCGTGCGLGSGGACRGGLGCGCGAGAAAAAAGAGAGCGGGLGCGGA  344
Db      933  SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGTGCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGS  992
Oy      345  GAAAAAGAGAGCGGGLGCGGCGGAGAA-----AAAAAAGAGVROGCGGGLGS--GAGARG  398
Db      993  GAGAGSGAGAGSGAGAGSGAGAGSGAGAGCGAAGTGCAGAGSGAGAGSGAGAGSGAGAGSGAGAGS  1052
Oy      399  ---GCGAGAAAAAAGAGAGCGGCTGGLGCGGCGGAGGGLGCGGAGAAAAVGAAGCGGCGGAGS  455
Db      1053  AGAGSGAGAGSGAGAGSGAGAGSGAAGTGCAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA  1112
Oy      456  SAAGAAASRLSSPQASRSRVSAVSNLVAAGPTMSAALSTISINYSQIGASRPLGSGDV  515
Db      1113  GAGGAGC-----AGSGAGAGSGAGAGSGAGAGSGAAGTGCAGAGSGAGAGSGAGAGSGAGAMPG-----  1159
Oy      516  LIQALLGHHH  526
Db      1160  RYQLSAGRYHY  1170

RESULT 15
US-08-175-155-48
: Sequence 48, Application US/08175155
: Patent No. 5641648
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
APPLICANT: Cirlsman, John W.
APPLICANT: Dorman, Mary A.
TITLE OF INVENTION: Methods for Preparing Synthetic
TITLE OF INVENTION: Repetitive DNA
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fleht, Honbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175,155
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertam I.
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-5/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1059 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-175-155-48

```

Best Local Similarity 45.48; Pred. No. 1.4e-63;
Matches 244; Conservative 33; Mismatches 225; Indels 35; Gaps 18;

QY 17 GYGGLGGGAGGCGGAGAAAAAAG -GAGGCGGGTGLGSGGAGRG -GGGAGAAAAAAGAGG 74

Db 524 GAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS -GAGAGSGAGAGSGAGAGSGAGAG 582

QY 75 GYGGLGS -QGAGRGGGLGGGCGAGAAAAAAG - -GVGGGGLGGGCGAGCGGAGAAAAAGAGG 130

.Db 583 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 642

QY 131 GYGGLGS - -QGAGRGGSGGCGAGAAAAAAGAGGCGGGTGLGS -QGAGRGLGGCGAGAA 186

Db 643 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 702

QY 187 AAAAAAG - -GAGGCGTGGC - -GGGAGCGGTGGLGS -QGAGRGLGGCGAGAAAAAG -G 238

Db 703 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 762

QY 239 AGGCGGLGGCGGAAAAAAGAGAGCGGTGTCGLGS -QGAGR -CGGAGAAAAAAGAGAGCGGTGG 296

Db 763 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 822

QY 297 QGAGGCGTGGGLGS -QGAGRGLGGCGGAGAAAAAAGAGGCGGLGGCGGAGAAAAAGAG 355

Db 823 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 882

QY 356 QGGLGGCGAGGAGAAAAAAGAGGVRGGTGGTGLGSOGAGRG -GGGAGAAAAAAGAGG 414

Db 883 SGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGS -GAGAGSGAGAGSGAGAGSGAGA 941

QY 415 G - - - -GTGGLGGCGGTGAGGLGGCGGAGAAAAAGAGCGGTGGTGGSGAGSAAASRLSSPQ 469

Db 942 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGA 1001

QY 470 ASSSVSSAVENVLASGPTNAAALSLPTLSNVQGLGASNGTGLSGCVLLQALLGHHH 526

Db 1002 GSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAGSGAGAG 1052

Search completed: July 3, 2001, 14:56:45
Job time: 535 sec

Query match 35.68; Score 962.5; DB 1; Length 1059;

Db	438	AAAAAAGGAGGGCGYGGGLGNQAGRGRCGGCAAAAAAGAGAGCGCGYGGGLGSGAGRCGGCGAGAA	497
Qy	337	-----GGLGGCGAGAAAAAGAGGCGGGLGCGAG	365
Db	498	AAAVAGAGGEGIRGAGCGAGCGYGGGLGSGGSGRCGLGGCGAGAAAAAGAGAGCGGGLGCGAG	557
Qy	366	QGAGAAAAAAGAGGVRGCGYGGGLGSGGAGRGCGAGAAAAAGAGAGCGCGGTGGGLGCGCV	425
Db	558	QGAG-----AAAAAAGCGVRGGGTGGGLGSGGAGRGCGAGAAAAAGAGAGCGGGTGGGLGCGCV	613
Qy	426	GAGGLGCGGAGAAAAVGAAGCGGCGYGGVSGCASAASAAARLSSPQASRYSAVSNLVASG	485
Db	614	GRGGLGCGGAGAAAAAGAGAGCGGCGYGGVSGCASAASAAARLSSPQASRLSSAVSNLVATG	673
Qy	486	PTNSAALSTTISNVYSQTGAS	506
Db	674	PTNSAALSTTISNVYSQTGAS	694

```

RESULT      2
A44112
spidroin 2, dragline silk fibroin - orb spider (Nephila clavipes) (fragment)
N:Alternate names: silk fibroin, dragline
C:Species: Nephila clavipes
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Mar-1998
C:Accession: A44112; S27824
R:Hilman, M.B.; Lewis, R.V.
J. Biol. Chem. 267, 19320-19324, 1992
A:Title: Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes
A:Reference number: A44112; MUID:92406876
A:Accession: A44112
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-627 <H12>
A:Cross-references: GB:M92913; NID:g159713; PTD:g159714
R:Hilman, M.B.; Lewis, R.V.
submitted to the EMBL Data Library, May 1992
A:Description: Isolation of a clone encoding a second dragline silk fibroin: Nephila clava
A:Reference number: S27824
A:Accession: S27824
A:Molecule type: mRNA
A:Residues: 19-627 <H12>
A:Cross-references: EMBL:M92913

```

Query Match	42.5%	Score 1148;	DB 2;	Length 627;
Best Local Similarity	51.1%	Pred. No. 8,6e-56;		
Matches 314;	Conservative 25;	Mismatches 139;	Indels 136;	Gaps
QY 17 GYG----GLGGQGAQ-----GAG-AAAAAAAAAGGAGGYG-----GLGSQAGR-- 56				
Db 3 GYRGQGGPGGTYGRGCGQGGSPSPGSAAAAAAAAAAAGPGTYGPQQGGPGGTYGPQQGGPGRY 62				
QY 57 ----GGQGAGAAAAAGAGAGGAGGGYGGGLGSQAGRGGLG-GQ----GAGAAAAAGV--- 103				
Db 63 PGQGGSPGPGSAAAAAAGSGCGQGGPGGYPGQQGGGGYGGQGGQGGSPGSGSAAAAASAE 122				
QY 104 -GGGGLGGQGGAGGAGAAAAAAGCAGGGGYG----GLGSQAGRGGSGSGCGG-AGAAAAA 157				
Db 123 SGQGGPGGYGYGQ-----QGEGGGYGGCGQGGPGGYYGPQQGGPGSGSAAAAAAA 171				
QY 158 GGAGQ---GGYG---GLGSQAGRGGLGQGAAGAAAAAAGCAGGCGYGGGLGQGAQGG 210				
Db 172 SGPGQGGCGGYPGPGQGGPGGYPGCGQGGSPSGSAAAAAASGPGQGGPGGTYGPQQGGP 231				
QY 211 GYGGLGSQAGRGGLGQGAAGAAAAAGCAGGGGLGGGAG-----AAAAAAGGAGQ 261				
Db 232 GY-----GPGQGLSGPGSAAAAAAGPGQGGPGGYGPGQGGSPSGSAAAAAAGP 285				
QY 262 GGYG----GLGSQAG-RGGQAG-AAAAAAGAGAGGCGYGGQGAAGGCGYGGGLSGAGRG 315				
Db 286 GGYGPGQGGPGGYPGQGGPGGAGSAAAAAAGCGQGGGLGYPGQGGPGG---YGPQGG 342				

QY 316 GLGGGCGAATAAAGAGCGGGLGGGAG-----AAAAAG--GAGGCG 358

Db 343 GPGGTGPGSASAAAAGGPGGQGGGCGPGGQCGSGCGSAAAAAAGPGGTGGCGG 402

QY 359 LGGGGAG-----GAGAAAAAAGGVRGGYGGGLGSGAG-----RCGGGAG--AAAA 407

Db 403 PGGYAPGQGGPSGSGSAAAAAAG--PGGYG--CGGPGGYAPCGQGGSGSGSAAAA 459

QY 408 AAGGAGGCGTC-----GLGGGV-----GAGLG-----GGGAGAAAAGVAGGCG 447

Db 460 AAAAAAGPGGYPPAQQGSPGPIAASASACPGGGGPGAGCGPGSAYAAASAGAGSAG 519

QY 448 YGCVGSGAASAAASRLSPQASRSYASVNLVAGPNTAALSTTISNVVSGTGSN 507

Db 520 YG---PGSQAASAAASRLSPDCAARVAASVNLVSSGPTSSAAASVSIASVSIQAGSN 575

QY 508 PGLSGCVLIAL 521

Db 576 PGLSGCVLIAL 589

```

RESULT 3
T31328
fibroin - Chinese oak silkworm
C:Species: Antherea pernyi (Chinese oak silkworm) #text_change 22-Oct-1999
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999
C:Accession: T31328
R:Sezutsu, H.; Tamura, T.; Yukihito, K.
submitted to the EMBL Data Library, August 1998
A:Description: Characterization of the full length fibroin gene of a wild silkworm, A
A:Reference number: Z20995
A:Accession: T31328
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2639 <SEZ>
A:Cross-references: EMBL:AF083334; NID:g3450882; PID:g3450883; PIDN:AC32606.1
C:Genetics:
A:Introns: 14/3

```

Query Match	35.9%	Score 968.5	DB 2	Length 2639
Best Local Similarity	34.7%	Pred. No. 1.5e-45		
Matches 287	Conservative 30	Mismatches 160	Indels 351	Gaps 26
QY 17	GYG-GLGGGAGACGAGCAAAAAAGAGAGCGGCGYGLGSGCAGRRGCGG-----AGAAAA	66		
Db 780	GYGWMDDGGYGGSDSAAAAAAAAAAAAAAAAAGSAGAGAGGCGYGMDDGYGSDSAAAAAAAAAAAA	839		
QY 67	AAAGAGCGGCGGLGSGQ-----GAGRGGLG-----OGACAA	97		
Db 840	ASGAGGSGGCGYGGYSDSAAAAAAAAAAAAAAAAAGGAGGCGYGMDDGYGSDSAAAAAAAAAA	899		
QY 98	AAAGVGCGGGLGCGAGGAGCAAAAAAGGA-----	127		
Db 900	AAAAGSGAGCGRRDGGYGGSSSAAAAAAAAAAASARRGHDSAGSAAAAAAAAAAAAASG	959		
QY 128	-----GCGGYG-----GLGSGGAGRG--GSGCGC-----	149		
Db 960	AGSGCGGCGWGGGCGYGSDSLAAAAAAAAAAAAAAAAAGCAGGAGCGYGMDDGYGSDSAAAAAA	1019		
QY 150	AGAAAAAGAGGAGCGGCGGLGSGQ-----GAGRGGLG-----GCA	183		
Db 1020	AAAAAAAAGARSGGCGYGGYSDSAAAAAAAAAAAAAAAAAGSAGGCGGCGYGMDDGYGSDSAA	1079		
QY 184	GAAAAA-----AGAGCGCGYG-----	200		
Db 1080	AAAAAAAAAAGSGAGCGRGGCGYGSSSAAAAAAAAAAAAAAAAARRAGHDPRAAGSAAAAAAAAAA	1139		
QY 201	-----GLGCGGAGCG--GYGGLGSGGA-----GRGGLGCG-----A	230		
Db 1140	AAASGAGGSGGCGYGMDDGGYGSDSAIAAAAAAAAAAAAAAAAAAGAGSGGCGYGGYSDSAAAAAA	1199		

```

QY 231 GAAAAAGCAGCGGLGCG-----QGAGAAAAAGCAGCGCTGCLGSCGAG 273
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1200 AAAAAAAGSGAGGAGGCGGWDGCGYSDSAAAAAAAAAAAAAASGAGGCGGCGYGGYSD 1259
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 274 RCGCGACGAAAAAAGCAGCGCGY-GGCGAGCGCGYG-----GLCSGCGARC-- 315
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1260 SAAAAAAGAAAAAGSGAGGAGGCGTGMCDGCTGSDSAAAAAAAAAAGSAGGRGDBG 1319
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 316 -----GLGCGGAG----- 323
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1320 GSGSSAAAAAARRAGHIDRAAGSAAAAAAAAAAAAAASGAGSGGCGYGMCDGCGY 1379
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 324 -----AAAAAAGCAGCGGLGCG-----AGAAAAAAGCAGCGGL-CCGCGAG 365
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1380 SDSAAAAAAGAAAAAAGSAGAGSGCGYCGYSDSAAAAAAAAAAAAAAGSAGAGGCGYGMG 1439
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 366 QG-----AGAAAAAAGAAAGCGYCGYGLGSCGAGRGCGCGAGAAAAAAGCAGCGCT 417
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1440 DCGYCGSDSAAAAAAAAAAGSAGAGSGCGYCGYSDSAAAAAAAAAAGSAGAGCGY 1499
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 418 GGLGCG-----GVGAGGLG-----QGAGAAAVGAGCG 447
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1500 GGYGSDSAAAAAAAAAAGSAGAGCGYCGYGMCDGCGYSAAAAAAAAAAAGSAGG 1559
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
QY 448 Y--GCVGSGAASAAASRLSSPQASSRVSSAVSNILVASCTPNSAALS 493
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :
Db 1560 RGDGCGYGGSSAAAAAARRAGHIDRAAGSAAAAAAAAAAS 1607
    ||||| : : : : : ||||| : ||||| : ||||| : ||||| :

```

RESULT 4
F70806
hypothetical glycine-rich protein RV3508 - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70806
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference numbers: A70500; M01D:9825987
A:Accession: F70806
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1901 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CA117745.1; PID:g292444
A:Experimental source: strain H37Rv
C:Genetics
A:Gene: RV3508
A:Superfamily: collagen alpha 1(IV) chain

Query Match	34.68;	Score 935;	DB 2;	Length 1901;
Best Local Similarity	37.08;	Pred. No. 7.6e-44;		
Matches 264;	Conservative 24;	Mismatches 220;	Indels 206;	Gaps 28;

[illegible]

OY	138	GYGGLGSGOG-----AGGGGYGGLGSOG-----	AGRGLLGGDQ-CAGAAAA	235
Dd	902	GAGGAGAGADNNFNNGGGGGAGGGGGGLGCASSTTSINANGAGGNGSTTGGGAGGACT	961	
OY	236	AGGAGOGGLGSGOGA-----GMAAAAAGCAGCGGGLGSOGA-----	-G	273
Dd	962	LGVVSGSGGTGGDGDDAGSggcgGFCGAAGKAggGNGRGDDGDGAGSLGLGLGFDDG	1021	
OY	274	RGGGAGAAAAAGA--GGGTYGGGACGGTGGLGSAG- RGLLGGQ -GAGAA-	326	
Dd	1022	OGGGGAGGSAcAGGInGAGGAGNG- GDGGGATGAAGLLDNdgvgvGdGGAAGGAGNGG	1080	
OY	327	-----AGGAGGCG-----GQGGLGCG-----AGAAAA--	350	
Dd	1081	NAGVGLTKAKDDGGAAGNGCNGAGGAGGAGDNNPngGGGAGGGGGGLGASSTTSIN	1140	
OY	351	-----AGGAGGGGLGSGO-----AGGAGAAAAAAAAGGVRGG	386	
Dd	1141	AMGAGCNCGTGCKGAGGAGAGTLGVGSGSGTGGDGDAGSGGGGFGCAAKKAGGGGNGG	1200	
OY	387	YGGGLSGOA-----GRGGGAGAAAAAGG-----AGGGGTGGLGGOGV----	425	
Dd	1201	VGGDGEBSGLGLCLSGFDGGGGGGAGGAGGInGAGAGAGGTGAGGDDGPATLI	1260	
OY	426	-----GAGGLGCGAGAAAAVA-----	-----GGGGYGVGSGSAASA	460
Dd	1261	GGPDGDDGGGGITGGDGNAGFCAGVPDGDGCGNAGFCAGVPDGGGITGGTGAGAGAGA	1320	
OY	461	AASRLSPoASSRSVAVSNLYASCPtNSALSttSNVVSQIGASNGLGCD	514	
Dd	1321	GADGDPSTIDGGGGAGGHGGGGKGLNSTGtLASAASDdGNGGAGGAGGNGCD	1374	

B70807
 Hypothetical glycine-rich protein RV3512 - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: B70807
 R:COLE, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Salston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:982595987
 A:Accession: B70807
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1079 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17749.1; PID:g2922
 A:Experimental source: strain H37Rv

C; Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match	34.3%	Score 927;	DB 2;	Length 1079;
Best Local Similarity	40.2%	Pred. No. 1.3e+3;		
Matches 238;	Conservative 14;	Mismatches 202;	Indels 138;	Gaps 23;
QY	2	ASMTGGQQMGRRIRNGYGLGSGQ-----AGQGAGAAAAAAGAGAGGGY-----	47	
DB	315	AGGAAGSNMGCTYGANGTGGDDGNGGAAAGATRAGSNMGAGCTGAGGNGCTGGRGSGGAGC	374	
QY	48	GLGSGAGRGGQGAGAAAAAGAG-----QGGYGLGSGQ-----AGRGGJGSGGA	94	
DB	375	DEIGVGGGKKGNGADGEVCGAGGAGSGSPNTPSGGNGGGGGGSGGAGGAAGGAGG	434	
QY	95	GAAAAAGCYGGCGGLGCGAGGAGGAGAAAAAGAGGCGGTGGGLGSGGAGRGSGGQ--GAGAA	153	
DB	435	GANGTAGNMGCGGAGAGCTG---GAGAAASSATNG--GSGGAGGTGTGCGG--GSGGAGGTGTGAGCT	489	

QY	154	AAAGAGAGCGGCGGGLGSGAGRGGLGCG-----ACAAAAAAGG--AGCGG	199
		:	
Db	490	GGAAGDGGCGCGGAGGAGCGCGAGAGAGTGTGNGNITGTAGTAAGGAAAGKGA	549
QY	200	GGGCGGAGGCGGCGGGLGSGG-----AGRGGLGCGG---A	230
Db	550	GGCGGTGGGTGGCGGAGGDDGAGGTGGDRTVGGCTVPAGSGGGGCGNMGCGAGGCGGADG	609
QY	231	GAAGAAAGAGAGCGGCGG-----CGAGAAAAAAGAGAGCGGGLGSGAGRGGO--GAGA	281
		: : : : : : :	
Db	610	GGGGDDGDAAGTGGAGNGNGNNGNNGTGTGAGNGGGGAGNGAGGAGGSGGGGTGCGNGAGG	669
QY	282	AAAAAG-----GAGCGGCGGCG--AGCGGTG--GLGSGAGRGGLGCGGAGAAAAAAA	331
Db	670	DACGAGNGNGNCTGCGNGCGGAGTGGAGNGAGTGGGNGNGSGGNGGAGNGGNSG	729
QY	332	GGAGCGGCGGCGGAGAAAAAG-----AGCGGCGGCGAGGAGCA--AAAAAAG	380
		:	
Db	730	TGSDGGGAGGNGAAGTGTGTGGDGLTGTGTTGGSSGGTGGGCGNGGAGDNTAMTTAAG	789
QY	381	GVRRGGYGGGLGSGAGRGGCGAGAAAAAAGAGAGCGGTGGLGCGV-----	425
Db	790	G--DGGNGCGDGGFGGGAAGGAGGLTAGANGTGGCGAGAGDGGNCAICGHPLTDDPGNG	847
QY	426	GAGGCGGAGAAAAVGA-----GGGTGGV-----SGAASASAA	462
		:	
Db	848	GTTGGCTGTGTGGAGTGTGGLGCTGTGGDGGNGCGTGGGEGVGGAGCTGGAA	899

RESULT 6
D70807
hypothetical glycine-rich protein RV3514 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70807
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajadurai, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: D70807
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1489 <COL>
A:Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PIDN:CAA17751.1; PID:g292444
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3514
A:Superfamily: collagen alpha 1(IV) chain

Query Match	33.8%	Score 912.5	DB 2	Length 1489
Best Local Similarity	39.1%	Pred. No. 1e-42		
Matches 263	Conservative 19	Mismatches 202	Indels 189	Gaps 34
Oy	2	ASMTGCGOQGRIRINGYGGIGGQAGQ--GACAAAAAAGGAQGGYGGIGGQAGAGGQ--	59	
	178	AGGAGGWLFG---VEGAGGCGAGGGGTGGAGGPGGLIMGGGAGGAGVGGAGGGGTGGAGGGA	234	
Oy	60	---GAGAAAAA--GG---AGGGYGGIGGQSQ-----AGRGGLGGQ--GACAAAAAAGG	103	
Db	235	ELLTFAGGCGGGGTGGPGATGCTGGHGGVGGDGMWLPAGAGAGAGGAGAGAGSDGA	294	
Oy	104	--GGGIGGQAGGQAGGAGA--AAAAAGAGAGGGYGGIGGQSQ-----AGRGGSQGG--	148	
Db	295	LGGTGTGTGTGAGGAGGCGGALLLILGAGGGGGLGAGAGGGCTGAGGDDVTLGGVGGTGGG	354	
Oy	149	-----GAGAAA--AAAGGAGGGGGYGGIGGQSQAGRGGLGCGGACAAAAAAGGAG	195	
Db	355	GVGGVAGLGGAGGAGGAGGQLTSASGAAGNAGVGGAGGGG--DDGAGAGGAGADDPGATGGTG	413	

OY	196	-QGGYGGJLGGG-----AGCGGYGGLSGCA----	-GRGJLGGG-----AG	231
Dd	414	FAGCAGAGGAGGSSGAGGTNGSGAGGDDGGAGAGAGADNP	TIGTGGDGTGAAG	473
OY	232	AAAAAGGAGGCGJLGG--GAGAAAAAAGAGGCGYGJLGSOGA-		275
Dd	474	AGGAGCANGTGTTGTMIGTTGNACGVGAGGCGBDGAGAGADNP	PGATGCTGFAGAG	533
OY	276	GQ-GAGAAAAAAGGAGGCGYGGCGAGGCGYTGJLSOGA----	-GRGJLGGG--AGAAAA	328
Dd	534	GAGGAGGSSGAGTNGSAGAGGTG-CGGAGAGAGAGADNP	TIGTGGDGTGAAG	592
OY	329	AAGAGCGCGJLGGQ-GAGAAAAAAGAGCGJLGGGQ-----		370
Dd	593	GAGGAGTGGTGGGMITGTTGNAGVCGAGGCGDGGAGAGADNP	PGATGCTGFAGAGCA	652
OY	371	AAAAAAAAAAAGV-----RGGYGGJLGSOGA----	-GRGCGAGAAAAAG	410
Dd	653	GKAGSSSAGGTTNSGSAGCTGRQSCTGAGAGADNP	TIGTGGCGCTGCAAGAGAG	712
OY	411	G-AGGCGTGGCL-----GGCGV-----		443
Dd	713	GAAAGTGGGMITGTTGNAGVCGAGCASSGAGCTNGSGAGGT	GDGGCGAGAGADNP	772
OY	444	--GGGCGVSSGASAAASRLSPQARSRYSSAVSNLVASPTN	SALSTISWYS	501
Dd	773	ICGTGDCDGTGGAAGAGAGG--		810
OY	502	QTGASNPLSGCD	514	
Dd	811	--GVGAGGGGAD	821	

RESULT 7
 E70806
 hypothetical glycine-rich protein RV3507 - *Mycobacterium tuberculosis* (strain H37RV)
 C:Species: *Mycobacterium tuberculosis*
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: E70806
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garner, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A>Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70806
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1381 <COL>
 A:Cross-references: GB:AL022022; GB:AL123456; NID:93261554; PIDN:CAA17744.1; PID:9292
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3507
 C:Superfamily: collagen alpha 1(IV) chain

[illegible]

QY 148 QGAGAAAAAGCA-----GCGGCGGLGSG-----AGRGGLGCG--A 183
 DB 321 QCGGCGGTGAGAGAPVLFHGGAGAGKGGGCTGCMGAGAGDGTTVIAAGTGGEGGTGAA 380
 QY 184 GAAAAAAGCA-----GCGGCGGLGSG-----GCGGY 212
 DB 381 GAGCAAGARALTLTSGGLAGVAGAGTGTGTGTGNGADAAAVFGANGDPFGAGKGN 440
 QY 213 GGLGSG-----AGRGGLGCG-----AGAAAAAG-AGGGLGSG--GAGAAA 254
 DB 441 GGIGGAATVGTGADGGTGGKGTGTGAGAGAGADAGSTGNPGKGGDGGIGGAGAGAGAA 500
 QY 255 A-----AGAGAGCGTGGGLGSGA--GRCG-----GAGAAAAA 286
 DB 501 TGNCGHAGNTGDDGDTGTGNGCGNGTGTGVCADTLNPTDPGAGEPAGAGAGAGAAA 560
 QY 287 GG-----AGGCGGCGG-----AGGCGGCGGLGSG-----AGRGGLGCG--A 324
 DB 561 GCGPGCTGTGTGNGCGNGCGNGCGNGCGNGCGNNGNNAAPVCGEGAGAGDGGAGAGAG 620
 QY 325 AAAAAAGAGAGCGGLGSG-----AGAA 347
 DB 621 AANCGTAGSGCTGTGVDGAGAGNGGGKAGTNSNPGVDEAGFSGAGAGNGVGGAA 680
 QY 348 AAAAGAGCGGGLGSGAGAGCA-----AAAAAAGCGVRGCGTGGGLGSGA 395
 DB 681 GANCGTGGSGGNGDAGAGGAGGAGNGIPCTGTETPAAGTCAKGGDGGDGGAGAGAGAG 740
 QY 396 GRCGCGAGAAAAAGAG-----OGGTGGLGCGGAGGLGSGGAGAA 438
 DB 741 GAGCGGAGAGAGAGAGAGNATVPGDGVKAPKPDAGAGSGGDKRG--GQGGSGGTG--GSG 798
 QY 439 AAVAGAGCGTGGV-----SGASAASAAASRLSSPOASRRVSAVS--NLVASGPTNSAA 491
 DB 799 APICGAGGTGGSGHAGKAGAGGAGTITVTPNGCGNAGDGGNGAGAGNGSGSD 858
 QY 492 LSSTSNVSOIGAS--NPLGSG 512
 DB 859 FCGNTTSGASGSGGNGAGTAG 881
 RESULT 8
 E70895
 hypothetical glycine-rich protein RV1087 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
 C:Accession: E70895
 R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:9825987
 A:Accession: E70895
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-767 <COL>
 A:Cross-references: GB:AL021897; GB:AL123456; NID:93256022; PIDN:CAA17203.1; PID:el25196
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV1087
 C:Superfamily: unassigned collagens

Query Match 32.6%; Score 880; DB 2; Length 767;
 Best Local Similarity 42.2%; Pred. No. 3,5e-41;
 Matches 245; Conservative 22; Mismatches 218; Indels 96; Gaps 26;

QY 1 MASTGGGQWRIRIRYGGGLGAGAGAGAAAAAAGAGGCGGGLG--SOGAGRGG 59
 DB 155 MAGNGC--AACTLGNCGFGCGGCPAAGAGNGAGAGLFGNGAGAGGLGAVPVGAG 213

QY 60 GAGAAAAAGAGCGGCGGLGSGAGRGGLGCGA--CAAAAGVGGGGLGSGAGAG 117
 DB 214 GAGAGGCGGAGLGHGAG--GAGGAGVAGAGCFEETIGAGAGGAGGAGAGAG 271
 QY 118 G-----AAAAAG--AGGCGGGLGSG--AGRGSG--GAGAAAAAGAGCGG 167
 DB 272 GGLXGDAAGAGDGGVAGAGTGTGLGNRAGAGAGAGAGAGAGAGAGAGAGAGAG 331
 QY 168 LGSAGRC-----GLGGGAGAAAAAAGAGAGG--YGLGSGAGAGG 213
 DB 332 TGG--GAGLGAQSYTFSSLSGLSGDDGAGAGAGAGAGAGTGTGMLY--GGAAGSGDG 388
 QY 214 GLGSGAGRGGLG-----GCGAAGAAAAAGAGCGGLGCGA-----GAAAAAG 258
 DB 389 GTGCG--GAGGAGVFLFSGSGGPGGCGVGGVGGAGAGAGLFGVGLGAGAGD 447
 QY 259 AGGCGGGLGSG-----AGRGGAGAAAAAGAG--GCGGAGGAG--GCGGCG 306
 DB 448 SCGGGFGGPGLAGLFGNPGNGVGGIGDAAAGAGAGAGAGAGAGAGAGAGAG 507
 QY 307 LGSAG--AGRGGLGCG--AGAAAAAAGAGAGCGGLGCG--GAGAAAAAAGAGAGCGGLGCG 362
 DB 508 SGDGGAAGRCGAGAGLGSAGGATNPAGNPGSGSVGIGGAGAGAGTALFGDGAAGAG 567
 QY 363 GAGGAGAAAAAAGAGAGCGGCGGLGSGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 568 GAAGGFGGISAATPAGSEAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 627
 QY 421 -GGG-----YAGGLGCGAGAGAAAG--AGGCGGCG--SGASAASAAASRLSSPOAS 471
 DB 628 PGGGAGDAGDGGAG 679
 QY 472 SRVSAVSNLVASGPTNSALSTISNVSOIGASNPGLSG 512
 DB 680 -----PGPTNGLNAGAGCGCGVGNATGPTLFG 708
 RESULT 9
 A70869
 hypothetical glycine-rich protein RV2490C - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70869
 R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:9825987
 A:Accession: A70869
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1660 <COL>
 A:Cross-references: GB:AL021246; GB:AL123456; NID:93261507; PIDN:CAA16067.1; PID:g279
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV2490C
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 32.2%; Score 868.5; DB 2; Length 1660;
 Best Local Similarity 41.5%; Pred. No. 2,7e-40;
 Matches 245; Conservative 22; Mismatches 175; Indels 149; Gaps 30;

QY 17 GYGLGCGAGAGAGAAAAAGAGCGGCGGLGSG-----GAGRGGAGAGAAAAAGAG 72
 DB 649 GKGAGGAGATGATGATGATVHSGNGCGKCGNGADATVAAANGKAGAGAGGLVBDG 708
 QY 73 GCGGCGGLGSGA-----GRCGLGCGGAGAAAAAG-- 102
 DB 709 AGDGGGAG 768

QY 103 -VGCGGGLGGCG-ACGGAG-AAAAAAGA-----GGGCTGGTGSQA--GRGSG 146
Db 769 ALGGGNGKRRDADQAGAGAGAGAGAGAGSVSDGGPBGCKGAGAGAGASGGGGKG 828

QY 147 GQGAGAAAAAAGAGAGGCGTGLGSQSAGRGGLGGQGAAGAAAAAAGCAGCGCTGLGGQG 206
Db 829 ASGADSAAEAVGAGAGKGGDGCGVGVG--GDGCPGCDG--GAGAAAAPAGQVSHGVGVGCDG 886

QY 207 -----AQCGTGGLSQA-----GRGLGG----QGAAAAAG-----GAGGG 243
Db 887 GLGAGAGGCGGGHGBDGGDGGDPGAGLGLGCDSENGTRRAASVDASDHGPBSGG 946

QY 244 LGGCGA-GAAAAAAGA-CGCGYG---LGSGA-GRGCGAGAAAAAGA----- 289
Db 947 NGGNNGNAQAQSAVAGAGAGNGDGMNGRVBDGAGNGDGGAAAGANGANGAPCSALA 1006

QY 290 -GQ--GCTGGG-----AQCGTGGLGSQAGRGLGGO-GAGAAAAAAGAGAGCGG 338
Db 1007 LGGPGNGOGDDAGQAGGAGAGAGAGAGGAGGVSQDGGAGNGAGNGCVAGASGGAGARG 1066

QY 339 LGG-----QGAAAAAAGAGAGCGGLGCGAGAGAGAAAAA-----GGVRGGGTGLG 391
Db 1067 AMGIDSTIGTTGAGGGGGDGGAGCGVGHGCGDGYGAAPSTGVSSHGTGV--GGDGLG 1124

QY 392 SQG--AGRGG-----GAGAAAAAAGAGAGGCTGLG-----QGV----- 425
Db 1125 GAGGCGAGAGMGIGITVTGAGAGAGAGNGCDPCAGCRGLGAGSGNGTSANAGVDAASKHP 1184

QY 426 ---GAGGCGGAGAAAAVG--AGCGYGG-----YSAGAAASAA 462
Db 1185 LTGGGGVGGCAKRAAAAGDGGDGGNMALFCDGAGCGDGDAGDTAAEA 1235

RESULT 10
S01820
glycine-rich cell wall protein 1.8 precursor - kidney bean
C:Species: *Phaseolus vulgaris* (Kidney bean)
C:Date: 30-Sep-1999 #sequence_revision 19-May-1994 #text_change 16-Jul-1999
R:Accession: S01820
R:Keller, B.; Sauer, N.; Lamb, C.J.
EMBO J. 7, 3625-3633, 1988
A:Title: Glycine-rich cell wall proteins in bean: gene structure and association of the
A:Reference number: S01820; MUID:89091109
A:Accession: S01820
A:Molecule type: DNA
A:Residues: 1-465 <KEL>
A:Cross-references: EMBL:X13596; NID:g21002; PIDN:CAA1932.1; PID:g21003
C:Comment: This protein is enriched in the cell wall fraction of young hypocotyls and on
tlys.
C:Comment: Much of the sequence consists of tandemly repeated 22-residue segments with
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8
C:Keywords: cell wall; structural protein; tandem repeat
#1-30/Domains: signal sequence #status predicted <SIG>
#31-465/Product: glycine-rich cell wall protein 1.8 #status predicted <MAT>

Query Match	32.1%	Score 866.5;	DB 1;	Length 465;
Best Local Similarity	47.88;	Pred. No. 1.2e-40;		
Matches 218; Conservative	8;	Mismatches 177;	Indels 53;	Gaps 19;

[illegible]

Db	202	GEHGGGGGGGGGAGGGGTGAGGEHGGAG--GGGGGAGAGGTGAGGGEH--GGAGGGGG	258
OY	253	AAAAAGAGGGTGGTGGTGGAGAGAAAAAGAGAGGGGGGGGGGGTGGTGGAG	312
Db	259	GGAGGGGTGAGGEH--GGAGGGGG--GAGGTGAGGGEHGGAGGGGGGAGGGTGA	316
OY	313	GRGLGGGGAGAAAAAAAGAGGGTGGAGAAAAAGAGGGT--GGGAGGGAGA	370
Db	317	GGGG--GGGGGAGGGTAAVGEHGGGGTGGGGGG--DGGGTGGEHGGGTGGGGG--	370
OY	371	AAAAAAAGCVRGGTGGTGGAGRRGGGAGAAAAA--GAGGGTGTGTL-----	420
Db	371	--AGGGTGTGGEHGGGTG--GGGGG--GGTACGGHGAATGGGCGGGGGGGGAG	425
OY	421	--GGGVTAGGTLGGGAGAAAAAAGAGGGGTGGVSG	454
Db	426	HGGGTGGAGGGGGGTGAGAAH--GGGTGGGGTGG	459

RESULT 11
A:70896
hypothetical glycine-rich protein Rv1091 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence,revision 17-Jul-1998 #text,change 24-Nov-1999
C:Accession: A70896
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
R.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajadream, W.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70896
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <COD>
A:Cross-references: GB:AL021897; GB:AL123456; NID:3256022; PIDN:CAA17207.1; PID:el22
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1091
C:Superfamily: unassigned collagens

Query Match	31.8%	Score 858	DB 2	Length 853
Best Local Similarity	42.8%	Pred. No. 5.9e-40		
Matches 232	Conservative 18	Mismatches 184	Indels 108	Gaps 26

QY 7 GQOMRIRIRVCGILGSGGAGCGAGAAAAAAGAGAGGCGGVLGSGGA-----GRGQGA 63
Db 292 GSGSGHALLMTGAGAGGNGSGGTGAGGGSTAGACAGNGAGAGGGGTGCLFENGAGGHHG 351
QY 64 AAA-----AAGGAGQGGYGLGSGGAGRGGLGQ-----GAGAAAAAGVGGGL 108
Db 352 AAAGNGLAAGNVSSGGCGAGGTGGAGGDD--GAGGAGMARTLWVGAGGAG--GDGA 408
QY 109 GGQGAGCGAGAAAAAGCA---GGGTYGGLGSG-----AGRGSGGQGA----- 150
Db 409 GGAGGCGSGGLGNNNGGAGGSDGSGRGGTGGAEGEAGAGLIVGTGHGDDGAGGAAVWG 468
QY 151 -GAAAAAAGGAGQGGYCGILGSGGAGRGGLGSGGAGAA-----AAAA 191
Db 469 GDDGAAAGTGLAGAGRGAG---GSGGSGDDGGGGAAGPAGMLFGDGCAGGAGAA 524
QY 192 GGAG--QGSGYGLGSGGAGQGGYCGILGSGGA-----GRGILGCGAGAAAAAGCA--- 239
Db 525 GGAGGQAGGGGCGNGNG--GNGGNGGNGKATGCMILYNGGAGCGCATAGAGAGAGNVG 583
QY 240 -----GGGGLGCG--GAGAAAAAAGAGQGGYGLG--SOGAG--RCGGAGAAAAAAGG 288
Db 584 STNGGSGTNGGILGTCGSSGAGAGNGLLTVGAGGHHGASGAGDRGAGGTGFTSSDG 643
QY 289 AGQGGYGGGAGQGGYCGILGSGGAGRGGLGCG--GAGAAAAAAGAGAGGCGILGCGGAGAA 347

Db 644 A--GGDGDG--GNGAGGTGGLFCGAGNGGPGGSGAANDIGNGAGNNG--GTFDNGCGN 699
OY 348 AAAAGGAGCGGLGCGAGAGCA-----GAAAAAAGVRRGGGTG--GLGSG-- 394
Db 700 GCGGGGAGSGGDDGCGAGNGAMLPGNGAGGCGCGKCGAGCGGLGCGSFGLPLGLNGSGD 759
OY 395 AGRGGCA--GAAAAAAGGAGCGGTGGLGCGGV--GAGCLGGGAGAAAVAGAGCGYGG 450
Db 760 GGDGCGNAPGCVLYGNGAGAGCGGSSGGTGGTGGATGAGAGKGGDGDAGLLDGDGNGNGG 819
OY 451 VG 452
Db 820 AG 821

RESULT 12
B70812
hypothetical glycine-rich protein Rv0834c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: B70812
R:Coic, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulistion, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: B70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-882 <COD>
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CA117640.1; PID:g291688
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv0834c
C:Superfamily: elastin

Query Match	31.88:	Score 857.5:	DB 2:	Length 882:
Best Local Similarity	35.28:	Pred. NO. 6.5e-40:		
Matches 247:	Conservative 20:	Mismatches 226:	Indels 209:	Gaps 29:
OY	5	TGGGQMGRIIRINGYGGLGCGAGCGAGAAAAAAGCA---	GCGYGGLGSGAGRGCGGA	61
	111			
	111			
	111			
Db	157	TGGAG-GAAGGLRHGCTGAG---CTGASGKGCGTGMKMSGAGAGSGSGGAGG		212
OY	62	-----GAAAAAGACGCGTGGGLGSG-----	AGRGLG---QCGAAAAAG-102	
	111			
	111			
	111			
Db	213	NALMGIGGNGAGAGAAAGVNGVGAGAGALVAIGAGAGCAATTTGGAGGAGS		272
OY	103	-----VGGGGLGCGGA-----GCGAGAAAAAGCA-----	GCGYGGLGSGAG	141
	111			
	111			
	111			
Db	273	NALGLFLGLGSGCGGSMMSGAGCAGCGSCGAASPFGLDIGTAGAGCHGAGTNG-G		331
OY	142	RGSGCGC-----AGAAAAAGAG-----	QCGY	165
	111			
	111			
	111			
Db	332	AGGAGAGAGSSGTVEFALDLNMGAGAGNGCAATTGTGGAGGTGGAFFAEPFTGFAAYGA		391
OY	166	GGLGSGAGRGGLGCGAGAAAAAAGCAAGCGGCGGLGCGAGCGGCGGLGSGAG-RGG		224
	111			
	111			
	111			
Db	392	GGLGGAATGAGCTGTTGGVGAAGFAALGVGAGAGAGGAATETGGTGGAGGLGGLLG		451
OY	225	LGGCGA-GAAAAAGACGCGGLGCGAGAAAAAGAG-----	QGGYGGLGSG-----	271
	111			
	111			
	111			
Db	452	AGGAGCGPCGAASAGCGSGHGTGCDALGLIGAGTGGVGVGAATPTCGNGAGGSGTGLL		511
OY	272	---AGRGCGAGAAAAAAGAGAGCGGCGGAGGCGGGLGSGAG-----	RGLGSGAGCA	324
	111			
	111			
	111			
Db	512	GCGVGAAGGHGGGASVGTGSGSGAGDGFEPVGAGNGGNACTGVGVNANGANGGSAATGA		571
OY	335	AAA---AAAAG---AGCGGLGQ-----GAGAA-----		347

D _b	572	LAAVGAGAACGADATSGTGTGGFCGACGASRCLIFALCGAGAGAAGDASTGVGGPGCGTGTT	631
O _y	348	-----AAAAGGAG--GGCGLCGGCA-----GGCAGAAAAAANAAGCVR	383
D _b	632	AASPFGIALTCGACAGCGAGTSCATGTGAGGDGYFEGEIVVLGTLFCGACAGGAATTDCGA	693
O _y	384	QCGTGTGCTGCGAG-----RCGCAGCAAAAAGAGAGCGGTGTGTCG-----QCYCA	427
D _b	692	TGGAGGCFGCAAGAGIANLTFGSVLTGGGAG--GAGGTATGTGGCGAGGGGGGLSSPVLLIGIT	750
O _y	438	GGLCGCGAGAAAAYAGCGGCGTGTGCGSAAASAASRLSPHASRVSAVSMLVAISGPT	487
D _b	751	GGAGDGCGGALGVLC--GMGGDDGGCGGEAVAGIIV-----CGAGCAGCAAPT	796
O _y	488	NSAALSTSTINVSQDIS-----ASNPELSCDYLI	517
D _b	797	GGCGAGCGNGDALGLIVGCGCGCAAGTCFCAANTGGANGDTTI	838

RESULT 13
 A:70812
 hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70812
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garber, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70812
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-749 <COL>
 A:Cross-references: GB:AL022004; GB:AL123456; NID:93261550; PIDN:CAAI7639.1; PID:g2931
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV0833
 C:Superfamily: elastin

Query Match	31.5%	Score 850;	DB 2;	Length 749;
Best Local Similarity	36.5%	Pred. No.1.4e-39;		
Matches 241;	Conservative 21;	Mismatches 223;	Indels 176;	Gaps 25;
QY	7	GQGMGRIRIRGTGGIGLGGQ-----GAGQAGCAAAAAAGCA-GGGGTGGIGSGGAGRG	57	
Db	20	GGPAGLIGGGAGGAGGGSAAVAVIGGAGAGCAALLFFGAGGAGGAGGSGGSGAAGGAGG	79	
QY	58	GQGAGAAAAAAGCAGGCGGGGIGLSGGAG-----RGIGGGGAGAGAA-----	98	
Db	80	AGGAGGLFPAISGSGSGFGGFPASTGTGTGAGGTGTGAGGLFASGIVGVTGGGAGSGSGTGGVGT	139	
QY	99	-AAGGV-----GGGGIGGGGAGGAGCAAAAAAGGAGGGGTGGIGSGGAGRGSGGGGAGAA	153	
Db	140	GGAGGLFPAISGSGGAGGSGSGTGGAGGTGAGGLFGAGGAGGIGGGGNNHTGGHGAGGAGSAG	199	
QY	154	AAAAAGAGGCGGGTGGIGSGSG--AGRGIGLGGQ-----GAGAAAAA-----AGG	193	
Db	200	LIALDGGAGCGAGCAATTTGTGAGAGGAGKAAALLFFSGGAGGSGGAAATGPTGDTGNSGGAGG	259	
QY	194	A-----GGCGGTGIGGGGGA-GGGGTGGIGSGGAGRG--GLGGQG--AGAAAAAGGAG	240	
Db	260	AGGKAGLLEFGSSGAGGSGGAGGAFANGSTGTGAGAGGAGGAGLIONGNGSGSGTSTVATGGAG	319	
QY	241	GGLGGGCGAGAA-----AAAAAGAGGGGTGGL-----	267	
Db	320	NGGAGGAGGAGGAGLIONGNGSGSGMGMDAPCGTGGVGGIGLGLLLGLDGANMAPSTNPLHTAQ	379	
QY	268	-----GSGCA-----GRGGGAGCAAAAAAGG	288	

Db 380 QQALAAVNAPIQAVTGRPLTNGANGANGSAPGSGHGWMLEGGGTGGSGVSGAGGDDG 439
OY 289 A-----GGGGYGGGAGGGYGLGSOGA-----GRGGLGGGAGAAAAAAGGA- 334
Db 440 AGGILTFAGAGGAGAGAVTGTGATGSSGAGGALLFPAGAGAGAGSSGIGCFRAAGAG 499
OY 335 GGGGLGGGAGAAAAAAGAG-OGGLGGGAGGAG--AAAAAAGAGVROGGYGL 390
Db 500 GPGAGGLFPNGGAGAGSGSVSGAGGEGAGGLFAGGAGAGAGGNNVVGAGGA 559
OY 391 GSGG-----AGRGGAGAGAAAAAAGAG-----QGCTGGLGGGAGAGGLGGGA 435
Db 560 GGVGGLFGAGAGAGSGGGSVAGDSGAGNAGLAPLAGAGGGGGGCPGTGAGGPGG 619
OY 436 GAAAVGA-GGGGYGAGV---SGASAAASAAASRL-SSPQASRVSSAVSNIVASGPTNSA 490
Db 620 DAGLLVSGGAGGAGGAGGFLTTGGPGAGDAGLLFGSGGAGAGAGSGRTDLGAGAGAGK 679
OY 491 A 491
Db 680 A 680

RESULT 14
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70917
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: E70917
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1329 <COL>
A:Cross-references: GB:295844; GB:AL123456; NID:93250713; PIDN:CAB09271.1; PID:g2131046
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1450c
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 31.4%; Score 847; DB 2; Length 1329;
Best Local Similarity 38.9%; Pred. No. 3.3e-39;
Matches 243; Conservative 16; Mismatches 196; Indels 170; Gaps 27;
OY 5 TGGGCGRTRIRYGL--GGGAGGAGAAAAAAGAGGGYGLGSOGAGRGOGA 61
Db 409 SGGKAGGNGAGAGAGLVNNGAGGAGNGAPGAPPSGGDPNGGGGAGAG- GKGGDS- 466
OY 62 GAAAAAAGAGGGYGLGSOGA-GRGGLGGGAGAAAAAGGGYGLGSOGAGOG-- 118
Db 467 GAAGGAGGAGGAGGKGGNGATGATGAGADGTDG--GKGGNGAGAGGGAGG 524
OY 119 -----AAAAAAGAGGGYGLGSOGA-GRGSGGGGAGAAAAAAGAGG 163
Db 525 GKKALAAATHQDSMGAGGAGGNGAGGAGGNGAKGTFDNGGDDVGGNGGSGRGIG 584
OY 164 GYGGGLGSOGA-----GRGGLGGGAGAAAAAAGGA-GGGAGGLGSOGA 207
Db 585 GAGGIGAGAGTGAADGACATPTTSGGNGGTGGGAGAAATYAGAGAGGAGGNGGLVNG- 643
OY 208 GGGYGLGSOG-----AGRGGLGQ--GAGAAAAAG-----AGOGGLGOGA 249
Db 644 GAGGKGGDDMAGYAGSSPTTAGESGTSQNGAGGAGAGRGCDPGDGTGAGAGNCA 703
OY 250 GAAAAAAGGA--GGGYGLGSOGAGRGOGA-----GAAA 284

Db 704 NGANAATPGAGGCGGCHGCPAAG-GNCGGCGPGGLAGNLPFGNGIGGYGSGGKGGAG 762
OY 285 AAG-----GAGGGYGGG-----AGGGYGLGS-----OGA-----GRG 315
Db 763 LAGDGNANGNFAFDGNDGSHGNGGNPCAGGGGSGAGSTPGAGAHGFTPTSGDG 822
OY 316 GLGGG-----AGAAAAAAGAGAGGGGLGGGGA----- 344
Db 823 GGGGNGNSQYVGGNGDGGGNGGNGSGAGTGNGRGGDGAFPGKSNATNPGENPNGN 882
OY 345 ---GAAAAAAGAG-OGGLGGGAGGAGAAAAAAGGVR-----OGGYGLG 391
Db 883 PGNGAGAGAGAGAGLNGNGAGAGNGGLGRTGNGAAGANVAVGAPGPGAGAGHGAG 942
OY 392 SGG--AGRGG-----GAGAAAAAAGAGGGGGLGGGOGV-----AGGLGGGAGAAA 440
Db 943 GNGGAGNGGGGAGVSDGAGAGAGGAGDAGDAGNGGAGAFAGGGGRRGGGNA 1002
OY 441 VGAGGGYGGYGGSGASASAAASRL 465
Db 1003 GNAGAGGPGGTGTAGKAGPAGSIL 1027

RESULT 15
F70963
hypothetical glycine-rich protein Rv2634c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 24-Nov-1999
C:Accession: F70963
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: F70963
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-778 <COL>
A:Cross-references: GB:280225; GB:AL123456; NID:93242265; PIDN:CAB02341.1; PID:e26639
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv2634c
C:Superfamily: unassigned collagens

Query Match 31.3%; Score 844; DB 2; Length 778;
Best Local Similarity 38.9%; Pred. No. 3.1e-39;
Matches 225; Conservative 24; Mismatches 204; Indels 126; Gaps 25;
OY 2 ASMTGGQ-----QMGRTIRYGLGGGAGG--AGAAAAAAGGA-GGGYGLGSOG 53
Db 207 ATLGGTGGVGTATGLISGGRFGGAGAAAGVTTGGVSGSGGVGFNGGFGGAGGLG 266
OY 54 ACRGGGAGAAAAAAGAGAGGGYGLGSOGA-----GRGGLG--QGAGAAAAAG-- 102
Db 267 AAGGVGGAASYCTGGGGVGDGAPGCGGAGPILLTGNGVGGGLGAGAAAGNGAGAGM 326
OY 103 -YGGGGLGGG-----AGGAGAAAAAAGGA-GGGYGLGSOGAG----- 141
Db 327 LLDGGAGGGGPAVAVAGVAGGAGAGNGNANWFSGSAGGOGGTGLAGTNGVNGSTA 386
OY 142 -----RGSGGAGAAAAAAGAGAGGGYGLGSOGAGRGGLGQ-CA 183
Db 387 NNTANGTDSNGNGNQNGGNGGPPAGVGAGV--GGGGLGSLDGNNGTGGKGA 444
OY 184 GAAAAAAGAGAGGGYGLG-----GGAGGGYGLGSOG-----AGRGGLGGGAGA 232
Db 445 GGTAGTGGAGAGGAGGAGGIGETDGSAGVATGEGGDDATGVDGAGGAGGGGGHNT 504
OY 233 AAAAGAGGGGLGGGAGAAAAAAGAGAGGGYGLGSOG-----AGRGGAG 280

Db	505	GVGDPAFGDGGIGGDDGNCALCAAAGNCGTGGAGNGRGHLLTGNGAGGAGGTGGTGGG	564
Qy	281	AAAAAGAGAGCGCTVGGCG-----AGCGGTGGLCSQ-----AGRGLLGCGAGAGAA	326
Db	565	GAGAGAGCV--GGAGGEBLTDGACTABRGCTGGLGCGTGTGGMGSGVGGNGGAAGS	622
Qy	337	AAAAAGCAGCGGLCGCGAGAAAAAAGAGCGGGLCGCGAGGAGAGAAAAAAGATVHOGG	386
Db	623	LIGLGGCGGAGCGVCGTG-----GIGGIGGAGGNGAGAGCTTTGGATTIGGGGGTGG	674
Qy	387	YGGGLCSQG--AGRGQ--GAGAA-----AAAAAGAGCGGT--GGTGGGCVGAGGLCGQ	433
Db	675	VGGAGGTGCTGGAGGTGGSGCAGCLIGMGAAGCTAGCTGGCGGLCGCGG--GNGGNGCT	733
Qy	434	GA-----GAAAIVGACGCTGCVGSGAASAAASRLSSP	468
Db	734	GATGGCGGDFALG--GNGAGAGAGSGPGGSSGTCGNMPP	771

Search completed: July 3, 2001, 14:59:45
Job time: 370 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:10:47 ; Search time 37.35 Seconds
(without alignments)
484.254 Million cell updates/sec

Title: US-09-490-291-8
Perfect score: 2700

Sequence: 1 MASMTGGCGMGRIRIRGCG.....GLSGCDVLQALGHHHHH 528

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues
Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2273	84.2	747	1 SPDL_NEPCL	P19837 nephila cla
2	1148	42.5	627	1 SPDL_NEPCL	P46847 nephila cla
3	952.5	35.3	5263	1 FBOH_BOMMO	P05780 bombyx mori
4	935	34.6	1901	1 YZ08_MYCTU	P10456 phaeolus v
5	866.5	32.1	465	1 GRP2_PHAUV	P71933 mycobacteri
6	844	31.3	778	1 YQ34_MYCTU	P06794 mycobacteri
7	839.5	31.1	914	1 WA22_MYCTU	P03211 epstein-bar
8	828.5	30.7	641	1 EBNI_EBV	P053810 mycobacteri
9	819.5	30.4	801	1 Y747_MYCTU	P56877 mycobacteri
10	807	29.9	957	1 Y278_MYCTU	P010637 mycobacteri
11	782.5	29.0	603	1 YD25_MYCTU	P09789 petunia hyb
12	736	27.3	384	1 GRP1_PETMY	P54330 mus musculu
13	729.5	27.0	860	1 ELS_MOUSE	P27443 arabidopsis
14	714.5	26.5	338	1 GRP_ARATH	P18185 mus musculu
15	697.5	25.8	481	1 LORI_MOUSE	P050630 ratius norv
16	684	25.3	543	1 YP91_MYCTU	P09372 ratius norv
17	680.5	25.2	384	1 ELS_RAT	P050630 mycobacteri
18	671.5	24.9	498	1 Y418_MYCTU	P05416 mycobacteri
19	612	22.7	463	1 Y418_MYCTU	P05416 mycobacteri
20	609.5	22.6	515	1 Y140_MYCTU	P05054 mycobacteri
21	608.5	22.5	434	1 YK98_MYCTU	P010707 mycobacteri
22	595	22.0	747	1 ELS_BOVIN	P040985 bos taurus
23	593.5	22.0	750	1 ELS_CHICK	P07916 gallus gall
24	536.5	19.9	730	1 ELS_HUMAN	P15502 homo sapien
25	526.5	19.5	252	1 GRP1_PHAUV	P10495 phaeolus v
26	516	19.1	1733	1 VNUA_PPVKA	P33485 pseudorabie
27	512.5	19.0	672	1 PHX5_MOUSE	P08389 mus musculu
28	470	17.4	2038	1 FSH_DROME	P13709 drosophila
29	455.5	16.9	966	1 FIB1_PETMA	P02674 petromyzon
30	454	16.8	316	1 LORI_HUMAN	P23490 homo sapien
31	439	16.1	671	1 CALI_RAT	P02454 ratius norv
32	434	16.0	1027	1 CAPF_RTEPA	P30754 riftinga pach
33	432	16.0	1156	1 GLH4_CABEL	O76743 caenorhabdi

34	429.5	15.9	1453	1	CAL1_CHICK	P02457 gallus gall
35	428.5	15.9	1466	1	CAL1_HUMAN	P02461 homo sapien
36	428	15.9	1464	1	CAL1_HUMAN	P02452 homo sapien
37	427	15.8	1453	1	CAL1_MOUSE	P11087 mus musculu
38	422	15.6	1459	1	CAL1_MOUSE	P28481 mus musculu
39	420	15.6	1460	1	CAL1_MOUSE	P28481 mus musculu
40	419	15.5	1373	1	CAL1_MOUSE	P01149 mus musculu
41	415.5	15.4	1366	1	CAL1_MOUSE	O06392 canis famli
42	415	15.4	1418	1	CAL1_MOUSE	P02458 homo sapien
43	411	15.2	1049	1	CAL1_BOVIN	P04258 bos taurus
44	409.5	15.2	1567	1	ICEN_XANCT	P18127 xanthomonas
45	409	15.1	1366	1	CAL1_HUMAN	P08123 homo sapien

ALIGNMENTS

RESULT	1	STANDARD:	PRT:	747 AA.
SPDL_NEPCL	SPDL_NEPCL			
AC	P19837			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	SPIDROIN 1 (DRAGLINE SILK FIBROIN 1) (FRAGMENT).			
OS	Nephila clavipes (orb spider).			
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;			
OC	Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.			
OX	NCBI_Taxid=6915;			
RN	[1]			
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.			
RX	MEDLINE=90384959; Pubmed=2402494;			
RA	Xu M., Lewis R.V.;			
RT	"Structure of a protein superfiber: spider dragline silk.";			
RT	Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).			
RN	[2]			
RP	SEQUENCE OF 653-747 FROM N.A.			
RX	MEDLINE=94165058; Pubmed=8120021;			
RA	Beckwith R., Arcidiacono S.;			
RT	"Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicenteparius (Araneidae).";			
RT	J. Biol. Chem. 269:6661-6663(1994).			
CC	-I- FUNCTION: THIS SPIDER MAJOR AMPULATE SILK POSSESSES UNIQUE CHARACTERISTICS OF STRENGTH AND ELASTICITY. IT MAY CONSIST OF PSEUDOCRYSTALLINE REGIONS OF ANTIPARALLEL BETA-SHEET INTERSPERSED WITH ELASTIC AMORPHOUS SEGMENTS.			
CC	-I- SUBUNIT: MAJOR SUBUNIT, WITH SPIDROIN 2, OF THE DRAGLINE SILK.			
CC	-I- SUBCELLULAR LOCATION: EXTRACELLULAR.			
CC	-----			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL: M37137; AAA29380.1; -			
DR	EMBL: U03848; AAB0212.1; -			
DR	PIR: A36068; A36068.			
KW	Silk; Repeat.			
FT	NON_TER	1	1	
FT	DOMAIN	1	25	25 X APPROXIMATE TANDEM REPEATS.
FT	REPEAT	1	38	
FT	REPEAT	26	36	
FT	REPEAT	39	66	
FT	REPEAT	67	96	
FT	REPEAT	97	130	
FT	REPEAT	131	158	
FT	REPEAT	159	191	
FT	REPEAT	192	204	
FT	REPEAT	205	235	

[illegible][illegible]

Db 2751 GAAGCAGAGSAGAGSGAAGSSGAGAGSAGAGSGAGAGCGATGAGTGGAGTG 2810

Oy 114 GCGAGAAAAAAGACGCCGTGGTGSQAGRCGSCGCAGAAAAAAGACGCCGTGGTGSQGA 173

Db 2811 GAGAGYCAAGAGSGAGSGAGAAGSGA-GAGSGAGAGSGAGAGSGAGAGSGAGAGS-GA 2868

Oy 174 GRGGLGGCGAGAAAAAAG-----GAGCG--GYGGLGGCGAGCGCGTGGTGGTGGAGRGGLGG 227

Db 2869 GAGSGAGSAGAGSAGAGAGCTGTGTGAGAGCACTGAGAGSAGAGSAGAGSAGAGSAG- 2927

Oy 228 QGAGAAAAAG---GAGCGGLGCGAGAAAAAAGAGAGCG-----GYGGLGSQAGARG 275

Db 2928 SGAGSGAGAGSAGAGSAGAGSAGAGAGYGTGAGYGTGAGYGTGAGAGAGAGAGSAGSG 2987

Oy 276 -----CGGAGAAAAAAGACGCCGTGGTGGCGCGCTGGTGS-----QGAGRGGLGCGAG 323

Db 2988 AGAGSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAG 3047

Oy 324 AAAAAAAAAAGCGCGCLGCGGAGAAAAAAGAGAGCG-----GLG-CGGAG 365

Db 3048 AGSGAGAGSAGAGSAGAGSAGAGSAGAGSAGAGSAGAGAGYGTGAGYGTGAGYGTGAG 3107

Oy 366 QGAGAAAAAAGAGYRQCGTGGTGGTGSQAGRG-QGAGAAAAAAGAGCGCTGGTGLG---- 421

Db 3108 AGSAGASGAGSAGAGSAGAGSAGAGSAGAGS-GAAGAGSAGAGSAGAGSAGAGSAGAGAGGA 3166

Oy 422 --GGGVGAG-----GLG-----GCGAGAAAAGAGCGCGYGTGGSASASAASAAAS 463

Db 3167 GTGAGYGTGAGYGTGAGAGYGTGAGAGAGYGTGAGAGSAGAGSAGAGSAGAGACTGAGAGSAGGA 3226

Oy 464 RLSSPQASSRVSAVSNLVASGPTNSAALSTISN 498

Db 3227 GYGAGAGAGSAGAGSAGAGAGAGTAGTCSSGCPGPYVAN 3261

RESULT 4

ID YZ08.MYCITU STANDARD; PRT: 1901 AA.

AC 053553;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE HYPOTHEtical PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.

GN RV3508 OR MTV023.15.

OS Mycobacterium tuberculosis.

OC Bacteria; Filicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxId=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-H37RV;

RA MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RA "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";

RT Nature 393:537-544(1998).

RL Nature 393:537-544(1998).

CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/

CC	or send email to license@slsb.slb.ch).					
DR	EMBL:	AL022022:	CAA17745.1:	-		
DR	HSPB:	P19972:	1KVD			
DR	Tuberculin:	Rv3508:	-			
DR	InterPro:	IPIR00084:	-			
DR	Pfam:	PF00934:	PE; 1.			
KW	Hypothetical protein; Repeat; Signal.					
FT	SIGNAL	1	30	POTENTIAL.		
FT	CHAIN	31	1901	HYPOTHEICAL PE-PGRS FAMILY PROTEIN		
FT				Rv3508		
SEQ	SEQUENCE	1901 AA:	147627 MM:	C7B1923D5D0146CD CRC64:		
<hr/>						
Query Match 34.6%; Score 935; DB 1; Length 1901;						
Best Local Similarity 37.0%; Pred. No. 1.le-35;						
Matches 264; Conservative 24; Mismatches 220; Indels 206; Gaps 28;						
OY	5	TCGG--QMGRIIRINGYGGLGSGAGCAAAA---AAGAG--QGCGTGLSGGACGCG	58			
Db	663	TGCAGVSVGNMIGTGCTGTGVGGAGAAAASSATGGAFFAGAGGGBGAGNGSV	722			
OY	59	QCAGAAAAAGAGGCGGYGGLSGSOGAGR-----GGTGGGAGAAAAAGVGCGGLGQ	111			
Db	723	GCTNMSGAGAGAGKRGRTGAGAGSSADNPFGAFRAGAGGTG--GAAGAGAGCATGTGT	781			
OY	112	GAGCGAGAAAAAAGAGCGGYGGLSGSAG-----RGSSCGGAGAAAAAGC-----	159			
Db	782	GGVVGATPSAGITGCGAGRGDGDASGLGLSLGFDPGGGCGGAGGSAGAGGTINGAG	841			
OY	160	-AGCGCYGCLSGSAG-----RGGLGQ--GAGAAA-----AAAGGAGCG	197			
Db	842	AGNGGDDGDATGAALGNDGVGGDGAGAGAANGCNAGVGLTKAKGDGAAGNGNG	901			
OY	198	GYGGLGSGQ-----AGDGGYGLSGSQ-----AGRGGLGQ--GAGAAA	235			
Db	902	GAGGAGGAGDNNFNCGGAGGCGGGLLGASTTSINANGAGAGNCGTGKKGAGGACT	961			
OY	236	AGGAGCGGLGCGA-----GAAAAAAGAGCGGTGGLGLSGA-----G	273			
Db	962	LGVSGSGGTGGDGDAGSGGCGFEGGAAGKAAGCGGNGRGDGDGCASGLGLSLGFPDG	1021			
OY	274	RGGAGAGAAAAAGA--GQGGYGGCGAGCGGYGGLSGSAG--RGGLGQ--GAGAAA----	326			
Db	1022	QGGGAGAGGSAGAGGTINGAGAGGNG--GDGDGATGAAGLGDNGGVGGDGCAGAGANG	1080			
OY	327	-----GAGAGGAG-----AAAAAGA-----GQGLGSGQ-----AGAAAA--	350			
Db	1081	NAGVGLTKAKGDDGCAAGCGNAGAGAGGAGDNNFNCGGCGGCGGCGGLGASTTSIN	1140			
OY	351	-----GAGAGCGGLGSGQ-----AGGAGAGAAAAAAGAVRGG	386			
Db	1141	ANGAGNGGTGCKGAGAGATLVGGSGGTGGDGDAGSGGCGFEGAAGKAGGGGNG	1200			
OY	387	YGLGSGQA-----GRGGGAGAAAAAGA-----AGDGTGGLGSGQV-----	425			
Db	1201	VGGDGEBASGLGLSLGSPFDGGGCGGAGGSAGAGGINGAGGAGCTGAGAGDGPATLI	1260			
OY	426	-----GAGGLGCGGAGAAAAVA-----GGGTGCVCSGSAASA	460			
Db	1261	GCPDGGDGGGDTGGDGNAGFGAIVPDDGDDGNAFGAGVPGGGTLGGTGAGAGAGA	1320			
OY	461	AASRLSPQASSRSVSNVLVASPTNSAALSSTISNVVSQIGANSGLSGCD	514			
Db	1321	GADGPSIDCGCGGAGGHGCGCGKGLNSTGLASAAASDGGNGAGAGANGCD	1374			
<hr/>						
RESULT	5					
GRP2_PHAU		STANDARD:	PRT:	465 AA.		
AC	P10496:					
DT	01-JUL-1989 (Rel. 11, Created)					
DT	01-JUL-1989 (Rel. 11, Last sequence update)					

DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1.8 PRECURSOR (GRP 1.8).
 OS Phaseolus vulgaris (Kidney Bean) (French bean).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
 OC Fabales; Fabaceae; Papilionoideae; Phaseolus.
 OX NCBI_TaxID=3885;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. TENDERGREEN; PubMed=3208742;
 RX MEDLINE=89091109; PubMed=3208742;
 RA Keller B., Sauer N., Lamb C.J.;
 RT Glycine-rich cell wall proteins in bean: gene structure and
 association of the protein with the vascular system.";
 RL EMBL J. 7:3625-3633(1988).
 CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
 CC -1- DOMAIN: THE REPEATED DOMAINS OF THE PROTEIN FORM A BETA-PLATED
 SHEET CONFIGURATION.
 CC -1- SIMILARITY: THE N-TERMINAL SIGNAL SEQUENCE OF ABOUT 30 AA SHOWS
 ABOUT 60% HOMOLOGY TO THAT OF THE GLYCINE-RICH CELL WALL
 STRUCTURAL PROTEIN GRP 1.0.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X13596; CAA31933.1; -.
 DR PIR: S01820; S01820.
 DR HSSP: P30129; 4DPV.
 KM Cell wall; Structural protein; Repeat; Signal.
 FT SIGNAL. 1 30 POTENTIAL.
 FT CHAIN 31 465 GLYCINE-RICH CELL WALL STRUCTURAL
 FT DOMAIN 33 465 GLY-RICH.
 FT FT 205 359 8 X 22 AA TANDEM REPEATS.
 SQ SEQUENCE 465 AA; 36683 MW; B5CA49B983B43607 CRC64;
 Query Match 32.1%; Score 866.5; DB 1; Length 465;
 Best Local Similarity 47.8%; Pred. No. 5.1e-33;
 Matches 218; Conservative 8; Mismatches 177; Indels 53; Gaps 19;
 QY 23 GGGGAGGAGAAAAAGAGGGGGLGSGAGAGGAGAAAAAGAGGGGGLGSG 82
 DB 33 GYGLGHGTGGGAGAGAGGGGGGGGGGAGAGGAGGGGGGGGGGGGGGG 92
 QY 83 GAGRGGLGSGAGAAAAAGGAGGGGGLGSGAGAGGAGAAAAAGAGGGGGLGSGAGR 142
 DB 93 GAGYGGGGGGGGGGGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 141
 QY 143 GGGGAGGAGAAAAAGAGAGGGGGLGSGAGAGG-----LGAGGAGAAAAAAG 193
 DB 142 GGGGGGAGGAGGAGAGGGGGLGSGAGAGGGGGGGGGGGGGGGGGGGGGGG 201
 QY 194 AGGGGGLGSGAGAGGAGGGGGLGSGAGAGGAGAAAAAGAGGGGGLGSGAGAA 252
 DB 202 GGGGGGGGGGGGGAGGAGGGGAGAGGGGGGGGGGGGGGGGGGGGGGGGG 258
 QY 253 AAAAGGAGGGGGLGSGAGAGGGGAGAAAAAGAGGGGAGGGGAGGGGGLGSGAG 312
 DB 259 GAGAGGAGAGGAGGAGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 316
 QY 313 GGGGAGGAGAAAAAGAGAGGGGGLGSGAGAGAAAAAGAGAGGGGGLGSGAG 370
 DB 317 GGGGAGGAGGGGAGGAGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 370
 QY 371 AAAAAAAGAGGGGAGGAGGGGAGGAGGAGGAGAAAAAGAGAGGGGAGGAGG 420

DB 371 -AGGAGTGGGAGGGG--GGGGG--GGYAGGDHGAAGYGGGGGGGGGGGG 425
 QY 421 -GGGAGGAGGAGGAGAAAAAGAGGGGAGGGG 454
 DB 426 HGGGAGGAGGAGGGGAGAGAAH--GGGAGGGGGGGGG 459
 RESULT 6
 Y034_MYCTU
 ID Y034_MYCTU STANDARD; PRT; 778 AA.
 AC P11933;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
 GN RV2634C OR MYCY441.04C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteriia; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hounsley T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Stalton J.E.,
 RA Taylor K., Whitehead S., Barrall B.G.;
 RT Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z80225; CAB02341.1; -.
 DR HSSP: P04002; 1ATF.
 DR Tuberculin; Rv2634C; -.
 DR InterPro: IPR000084; -.
 DR InterPro: IPR001899; -.
 DR Pfam: PF00934; PE; 1.
 KM Hypothetical protein.
 SQ SEQUENCE 778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;
 Query Match 31.3%; Score 844; DB 1; Length 778;
 Best Local Similarity 38.9%; Pred. No. 7.2e-33;
 Matches 225; Conservative 24; Mismatches 204; Indels 126; Gaps 25;
 QY 2 ASMTGGQ-----QMGRIIRIRYGLGAGAGG--AGAAAAAGAG--GGGAGGAG 53
 DB 207 ATLVGGTGGVGGATGLGSGGFGGAGGAAAGVGTGGGGGGGGGGGGGGGG 266
 QY 54 AGRGGAGGAGAAAAAGAGAGGGGGLGSGAG-----GRGLG--GGAGAAAAAG-- 102
 DB 267 AAGVGGAASFTGTGGGGGGGGGGGAGGAGPLLLGNGVGLGAGAAAGAGAGAG 326
 QY 103 -VGGGGLGAGG-----AGGAGAAAAAGAG--GGGAGGGLGSGAG----- 141
 DB 327 LLDGAGAGGGGPAVAGVGLGMPGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 386
 QY 142 -----RGGGGGAGAAAAAGAGAGGGGGLGSGAGAGGAGGAGGAGGAG 183

OY	17	GYGGLGGGCGAGCGCAAAAAAGGA--GCGGTGCTGCG--GAGRCGGCGCAAAAAAGGAG	72
Db	137	GNCGAGGSGA--PGAGACGNGCAGAGCTFGSCGAGCASTDYACGACGACGAGNAGLFGAAG	195
OY	73	QCGYTGIGLSQGA--GNGGLGCGGA--GAAAAAGVGGGGL--GGGAGCGGAGAAAAAGG	126
Db	196	VGGTGGGNSNGATGCGAGGAGAGAGCTFGACRGKSGGSGMLTGAGAGGAGNAGTLATGDBG	255
OY	127	AG-----QGCTGIGLSQGA-----GNGSSGCG--ACAAAAAG-----	158
Db	256	AGGTGASRSRGGGCGAGGAGCDAGMFFGGGCGAGGAGISKSYGDSAAAGAGAPGLIGNG	315
OY	159	-----GAGGGYGGIGLSGCG-----AGKGGIGCGGA-----	183
Db	316	GNNGNGCAGSTGGGDDGPGGAGCTGVLICNGKNGSGSGTGATLGRKAGTGTGVLGIDGF	375
OY	184	-----GAAAAAGAGG-----GAAAAAGAGG-----	195
Db	376	TAPASTPLHTLQDDVYINMWNDFQTLGRPLINGCANPTPGTGADGAGGMLFNGNG	435
OY	196	-----QGCTGIGLSQGA--GCGGTGCTLSQGA--GNGTLCGGAGAGAAAAAGAGCGG	243
Db	436	GCGTTCGVMNGAGAGAGAGCGCTLFTGTGTGGSGGPGATGLGCTGAG--GALLLFGSGGAG	494
OY	244	LCGGCA-----GAAAAAGAGGGGGYGGIGLSGCGAGGGGCGAGAAAAAGAGCGGGYG	295
Db	495	SGGGAAGVGNCGAGAGGNNGALLGAAGAGGAGAGAVG--GNGGAGNGGCLFANGAGGPGCF	553
OY	296	GCGAGCGGTGIGLSQGA--GNGTLCGGCGAGAAAAAGAGAGCGGL---GGGCGAGAAA	349
Db	554	GSPAGAGCTGAGAGNGGTLFAGCTGCGAGGGSTLAGGAGGAGNGGTLFAGCTGCGAGSHT	613
OY	350	AAG--GAGGGGLGCG-----GAGCGAGAAAAAGGVRGGGTGGLGCGAGGCGGA	402
Db	614	AAGVSGAGGAGCGAGGCLLSTLGASGAGAGSGSGSLTAAVY--GGTGGAGCTLFGSGGAGG	671
OY	403	GAAAAAGAGGCGGTGGLGCGGAGGAGGLGCGGCGAGAAAAAGAGCGGYCG---YGSASAA	458
Db	672	SGGFSNNGAGAGGAGGADACLTVGSSGAGAGASATGCAATGGGGAGGAGSAGFAGGDBG	731
OY	459	SAASRLSSPPASSRVSAAVSNLVAESPPTSAALSSTISNVVSOIGASNPGLSGCDVLIQ	518
Db	732	AGCATGHLGSLGAHHIGCKGCGVGSANVLINGNGNGNGNGNSGNAKSSGA--PGRSAGAGAGG	789
OY	519	ALLGHH 524	
Db	790	LLLGEN 795	
RESULT	10		
Y278_MYCTU	Y278_MYCTU	STANDARD;	PRT; 957 AA.
AC	P56877;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	HYPOTHEtical PE-PCRS FAMILY PROTEIN RV0278C PRECURSOR.		
GN	RV0278C OR MYV035.06C.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;		
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RA	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV.		
RX	MEDLINE=96295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parthill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F.,		
RA	Badcock R., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Sgares R., Sulston J.E.		

```

RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393.537-544(1998)
CC CC
CC -1 SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AL021930; CAA17353.1; -.
CC DR Tuberculins: RV0278c; -.
CC DR InterPro: IPR000084; -.
CC DR Pfam: PF00934; PE; 1.
CC KW Hypothetical protein; Repeat; Signal.
CC FT SIGNAL 1 30 POTENTIAL.
CC CHAIN 31 957 HYPOTHETICAL PE-DGRS FAMILY PROTEIN
CC FT RV0278C.
CC SEQUENCE 957 AA; 81905 MM; 71EBABDA17FDA47C CRC64;

```

	Query Match	29.98%	Score 807:	DB 1:	Length 957:
	Best Local Similarity	36.08%	Pred. No. 3.7e-30:		
	Matches 219:	Conservative 21:	Mismatches 195:	Indels 174:	Gaps 22:
OY	17 GYGGLGGGAGAGCGAAGAAAAA--GGAGQGGYGGGLGSGAGRGGGGAGAAAAAGAGAQG 74				
DB	265 GYGGGGGAGSSGGAGAGAGAGLPGAGTGGHGRFADSPG-GVGGAGAGGLFGAGGEG 323				
OY	75 GYGGLGSGAGRGGLGGGCA---GAAAAAGVG-----GGGLGGGAGGCGAGAAAA 123				
DB	324 GSGGSLTVAGGGCGAGAGNAGMLALGAAAGAGGICGGDGGTLTAGGIG--GAGAGGNAHL 381				
OY	124 AGGAGQGGYGGGLGSGAGARGGSGCG---CAGAAAAAGGAGCGGYYGGLGSGGA----- 173				
DB	382 FGSGGSGGAGGGGEPADGGGCGPGGAGTGVFSFG-GAGNGGGGGPAGIGAGGTGGILI 440				
OY	174 GRGGLGGGAGAAAAAGG-----AGGGYGGGLGSGAGAGCGYGGT----- 215				
DB	441 GNGGNGMGASAVTGGNGGIGGTGVLIGNGGSSGGTGAAGKAGVGVSGLLGLDFN 500				
OY	216 -----GSGGAGR-GGLGG----- 227				
DB	501 APASTPLHTLQONVNLNVNBPFTLTGRPLTGNANGAPGCGADGAGGWLFGNANGT 560				
OY	228 QGAGAAAAAGC--AGGGGLGGGA-GAAAAAGGA-----GGGTVGGLSQCGAGRG 276				
DB	561 PGTGAAGAGGWLFGNGNGGCHGANTATATAGGAGAGGILFTGNGNGTGITATGAG 620				
OY	277 -----QGAGAAAAAGGAGCGYGGGAGGCGGYGGGLGSGGA-----GRGGLGGGAGAAA 326				
DB	621 TGGAGGAGGVSLTIGSGGTGGNGNSIGVAGTGGAGRGGRDAGLLFGAAGTGGHGAAGV 680				
OY	327 AAAAAGGAGGGL-----GGGAGAAAAAGGAGGGL-----GGGGA 364				
DB	681 PAGVGGAGNGGLFANGGAGGAGGTNAAGNGNGNGLFTGTGGAGTFFGAGNGNGG 740				
OY	365 GCGAGAAAAAAGVAGYRGYG---GLGSGAGRGGCGAGAAAAAGAG----- 413				
DB	741 LFGAGGTGAAGSGSGITTTGGGCHGNAGLLSTLGAAGGAGSGGASSTLAGAGGTGGNG 800				
OY	414 -----GGGTGGLGGG---VGAGGLGGGCGGAAGAAVAGAGG 446				
DB	801 ALLFFRGAGAGHGAALITSIOGAGAGAGNGGLLFGSAGAGAGGSGANALAGTG 860				
OY	447 GYGGVSGGA 455				
DB	861 GTGGDGGHA 869				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 3, 2001, 15:05:35 ; Search time 95.41 Seconds
(Without alignments)
732.177 Million cell updates/sec

Title: US-09-490-291-8
2700

Sequence: 1 MASMTGGQGMGRIRRGYCG.....GLSGCDVLQALLGHHNNHH 528

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL.16:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2555	94.6	544	5	046171	046171 nephila cla
2	2313	85.7	617	5	046172	046172 nephila cla
3	1397	51.7	988	5	017434	017434 nephila cla
4	1099	40.7	636	5	016987	016987 araneus dia
5	968.5	35.9	2639	5	076786	076786 araneus p
6	935	34.6	1884	5	09NMW2	09NMW2 nephila mad
7	927	34.3	1079	2	053557	053557 mycobacteri
8	921	34.1	871	5	044358	044358 nephila cla
9	912.5	33.8	1489	2	053559	053559 mycobacteri
10	900	33.3	2249	5	09NMW4	09NMW4 nephila cla
11	892	33.0	1381	2	053552	053552 mycobacteri
12	890.5	33.0	738	5	002402	002402 pinctada fu
13	880	32.6	767	2	053435	053435 mycobacteri
14	877.5	32.5	1468	5	09GUB5	09GUB5 galliera me
15	868.5	32.2	1660	2	053215	053215 mycobacteri
16	862	31.9	907	5	044359	044359 nephila cla
17	858	31.8	853	2	053439	053439 mycobacteri
18	857.5	31.8	882	2	053845	053845 mycobacteri
19	850	31.5	749	2	053844	053844 mycobacteri

20	847	31.4	626	5	09NMW1	09NMW1 nephila mad
21	847	31.4	1329	2	006810	006810 mycobacteri
22	843.5	31.2	714	2	053556	053556 mycobacteri
23	836	31.0	1538	2	053395	053395 mycobacteri
24	833	30.9	923	2	053890	053890 mycobacteri
25	831	30.8	1306	2	053775	053775 mycobacteri
26	827.5	30.6	741	2	006808	006808 mycobacteri
27	820.5	30.4	484	2	053394	053394 mycobacteri
28	818	30.3	667	2	053415	053415 mycobacteri
29	813	30.1	1011	2	P94986	P94986 mycobacteri
30	812.5	30.1	783	2	053809	053809 mycobacteri
31	810.5	30.0	610	5	Q9V5V8	Q9V5V8 drosophila
32	810	30.0	731	2	050415	050415 mycobacteri
33	801.5	29.7	410	5	016988	016988 araneus dia
34	796.5	29.5	837	2	053684	053684 mycobacteri
35	793	29.4	576	2	P71664	P71664 mycobacteri
36	783.5	29.0	408	10	Q43688	Q43688 vigna ungu
37	781.5	28.9	694	2	053212	053212 mycobacteri
38	775.5	28.7	390	10	Q9M3Y2	Q9M3Y2 triticum ae
39	767	28.4	294	5	016986	016986 araneus dia
40	764.5	28.3	606	2	053884	053884 mycobacteri
41	758.5	28.1	1729	5	Q9U617	Q9U617 drosophila
42	748.5	27.7	584	2	053575	053575 mycobacteri
43	747.5	27.7	618	2	006801	006801 mycobacteri
44	747	27.7	615	2	005806	005806 mycobacteri
45	742.5	27.5	639	2	053952	053952 mycobacteri

ALIGNMENTS

RESULT 1
ID 046171 PRELIMINARY; PRT; 544 AA.
AC 046171;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SPIDROIN 1 (FRAGMENT).
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384959; PubMed=2402494;
RA Xu M., Lewis R.V.;
RT "Structure of a protein superfiber: spider dragline silk.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7120-7124(1990).
RN [2]
RP SEQUENCE OF 449-544 FROM N.A.
RX MEDLINE=98148687; PubMed=9487707;
RA Arcidiacono S., Mello C., Kaplan D., Cheley S., Bayley H.;
RT "Purification and characterization of recombinant spider silk
expressed in Escherichia coli.";
RL Appl. Microbiol. Biotechnol. 49:31-38(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX Beckwith R., Arcidiacono S., Store R.;
RT Insect Biochem. Mol. Biol. 0:0-0(1998).
RN [4]
RP SEQUENCE FROM N.A.
RX Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20329; AAC38957.1; -.
FT NON TER 1 1
SQ SEQUENCE 544 AA; 44107 MW; CC611B3551945615 CRC64;

Query Match 94.6%; Score 2555; DB 5; Length 544;
Best Local Similarity 99.6%; Pred. No. 5,86-143;
Matches 503; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	17	GYGGLGGCGAGAGGAGAAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTG	76
Db	1	GYGGGLGGGAGAGCGAGAAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTG	60
QY	77	GGTGGGAGRCGCTGCGAGAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTG	136
Db	61	GGTGGGAGRCGCTGCGAGAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTG	120
QY	137	SGAGAGRCGCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAGAAAAAAGAGAG	196
Db	121	SGAGAGRCGCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAGAAAAAAGAGAG	180
QY	197	GGTGGGAGRCGCGCGAGAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAGAAAAAAGAGAG	256
Db	181	GGTGGGAGRCGCGCGAGAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAGAAAAAAGAGAG	240
QY	257	GGAGAGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAG	316
Db	241	GGAGAGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCTGCGAGRCGCGAG	300
QY	317	LGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAG	376
Db	301	LGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAG	360
QY	377	AAAGAGRCGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGRCGCGAG	436
Db	361	AAAGAGRCGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGRCGCGAG	420
QY	437	AAAAGAGAGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGRCGCGAG	496
Db	421	AAAAGAGAGCGGCTGCGAGRCGCGAGAAAAAAGAGAGCGGCTGCTGCGAGRCGCGAGRCGCGAG	480
QY	497	SNVYSGTGAASNPGLSGCDVLIQALL 521	
Db	481	SNVYSGTGAASNPGLSGCDVLIQALL 505	
RESULT 2			
Q66172	1	PRELIMINARY: PRT: 617 AA.	
AC	046172:		
DT	01-JUN-1998 (TREMBLrel. 06, Created)		
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	DRAGONLINE SILK PROTEIN SPIDROIN 1 (FRAGMENT).		
OS	Nephila clavipes (orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.		
RN	NCBI_TaxId=6915;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	MEDLINE=94165058; PubMed=8120021;		
RA	Beckwith R., Arcidiacono S.;		
RT	"Sequence conservation in the C-terminal region of spider silk		
RT	proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus		
RT	bicentenarius (Araneidae)."		
RL	J. Biol. Chem. 269:6661-6663(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Beckwith R., Arcidiacono S., Stote R.;		
RA	Insect Biochem. Mol. Biol. 0:0-0(1998).		
DR	EMBL: U37520; AAC04504.1; ..		
FT	NON_TER 1 617		
FT	NON_TER 617 1		
SEQ	SEQUENCE 617 AA: 49410 MW: 1EFOCE9269A832E2 CR664;		

Query Match	85.7%	Score 2313;	DB 5;	Length 617;
Best Local Similarity	80.1%	Pred. NO. 1e-128;		
Matches 482;	Conservative 1;	Mismatches 5;	Indels 114;	Gaps 7;
<p>07 34 AAAAAGCAGCGCGTGGTGGCAGAGGCGAGAAAAAAAAGCAGCGGTGGTGGCAGAGCGTGGCG 93</p> <p> </p>				

Db	1	AAAAAGACAGCGGTGGLGSGGAGRGGGAGAAAAAAGAGCGCGGTGGLGSGGAGRGGGGLGCGG	60
QY	94	AG--AAAAAGGVTGGGGLGSGGAGGAGCAAAAAAGAGACGGGTGGLGSGGAGRGGSGGCGAG	151
Db	61	AGAAAAAGAGGVTGGGGLGSGGAGGAGCAAAAAAGAGACGGGTGGLGSGGAGRGGSGGCGAG	120
QY	152	AAAAAAGGAGCGGVTGGGLGSGGAGRGGGLGCGGAGAAAAAAGAGACGGGTGGLGSGGAGCGG	211
Db	121	AAAAAAGGAGCGGVTGGGLGSGGAGRGGGLGCGGAGAAAAAAGAGACGGGTGGLGSGGAGCGG	180
QY	212	YGGIGSGGAGRGGGLGCGGAG--AAAAAGAGCGGGLG---GGAGAAAAAAGGAGCGGTG	265
Db	181	YGGIGSGGAGRGGGLGCGGAGAAAAAAGAGCGGGLGCGGAGCGGAGAAAAAAGGAGCGGTG	240
QY	266	GLGSGGAGRGGGAGCAAAAAAGAGACGGGTGCGGAGCGGVTGGGLGSGGAGRGGGLGCGGAG	325
Db	241	GLGSGGAGRGGGAGCAAAAAAGAGACGGGTGCGGAGCGGVTGGGLGSGGAGRGGGLGCGGAG	300
QY	326	AAAAAG-----AAAAAG-----352	
Db	301	AAAAAGAGCGGGLGCGGAGCAAAAAAAGAGACGGGTGGLGCGGAGCGGAGAAAAAGGA	360
QY	333	-----GAGCGGLGCGGAGCAAAAAAGAGCG---GGLGCGGAG-----364	
Db	361	GCGGTGGLGSGGAGRGGGLGCGGAGCAAAAAAGAGCGGVTGGGLGCGGAGCGGVTGGGLGSGG	420
QY	365	-----CGGAGAAAA-----AAAAAGGVRGGGTGGLGSGGAGRGG	399
Db	421	RGGLGCGGAGCAAAAAAGAGCGGGLGCGGAGCGGAGCAAAAAAGGVRGGGTGGLGSGGAGRGG	480
QY	400	QGAGAAAAAAGAGACGGGTGGLGSGGAGGGLGCGGAGCAAAAAAGAGACGGGTGGLGSGGAG	459
Db	481	QGAGAAAAAAGAGACGGGTGGLGSGGAGGGLGCGGAGCAAAAAAGAGACGGGTGGLGSGGAG	540
QY	460	AAASRLSPQASRVSSAVSNLVAASPTSAALSTISNVVSIQIASNPGLSGCDVLIQA	519
Db	541	AAASRLSPQASRVSSAVSNLVAASPTSAALSTISNVVSIQIASNPGLSGCDVLIQA	600
QY	520	LL 521	
Db	601	LL 602	
RESULT 3			
ID	017434	PRELIMINARY;	PRT: 988 AA.
AC	017434		
DT	01-JAN-1998 (TREMblrel. 05, Created)		
DT	01-JAN-1998 (TREMblrel. 05, last sequence update)		
DT	01-OCT-2000 (TREMblrel. 15, last annotation update)		
DE	MINOR AMPULPATE SILK PROTEIN MISPL (FRAGMENT).		
OS	Nephila clavipes (orb spider).		
OC	Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;		
OC	Araneomorphae; Entelegynae; Araneolidae; Tetragnathidae; Nephila.		
OX	NCBI_TaxID=6915;		
RP	(1)		
RA	SEQUENCE FROM N.A.		
RX	MEDLINE=98200471; Pubmed=9541398;		
RT	Colgin M.A., Lewis R.V.;		
RT	"Spider minor ampulpate silk proteins contain new repetitive sequences		
RT	and highly conserved non-silk-like 'spacer regions'."		
RL	Protein Sci. 7:667-672(1998).		
DR	EMBL; AF027735; AAC14589.1; "		
DR	InterPro: IPR000817; "		
DR	PRINTS: PR00341; PRION.		
DR	NON_TER 1		
SQ	SEQUENCE 988 AA: 79082 MW: 461E03DF53F7085D CRC64:		

Query Match	51.7%;	Score 1397;	DB 5;	length 988;
Best Local Similarity	47.7%;	Pred. No. 7.8e-75;		
Matches 345;	Conservative 23;	Mismatches 144;	Indels 212;	Gaps 22;

01-JUN-1998 (TREMBLrel. 06, Created)
01-JUN-1998 (TREMBLrel. 06, Last sequence update)
01-MAR-2001 (TREMBLrel. 16, Last annotation update)
PGRS-FAMILY PROTEIN (FRAGMENT).
GN RV3512 OR MTW023.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RX MEDLINE=98295987; PubMed=9634230;
RC STRAIN-H37RV;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: AL022022; CAA17749.1; -
DR TuberculList: RV3512; -
DR InterPro: IPR002202; -
DR PROSITE: PS00318; HMG_COA_REDUCTASE_2; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 81163 MW; A79718CCDB7AB97D CRC64;

Query Match 34.3%; Score 927; DB 2; Length 1079;
Best Local Similarity 40.2%; Pred. No. 3.1e-47;
Matches 238; Conservative 14; Mismatches 202; Indels 138; Gaps 23;

2 ASMTGGGQCGRRIRNGYGGGCGG-----AGCGAGAAAAAGAGCGGCGY----- 47
315 AGGAAGSNGGTGANTGCTGGDGGGAAGATAGSNGAGTGSAGNGGTGGGSGGAGG 374
48 -GLGSGAGCGGAGAGAAAAAGAG-----OGYGGGLGSGG-----AGRGGLGCGGA 94
375 DGTGGYGGKGGKNGADPEVGGAGAGSGSPNTSPGNGGCGGCGGAGAGAGAGG 434
95 GAAAAAGVGGGGLGSGGAGAGAAAAAGAGCGGCGGLGSGAGRGSGGQ--GAGAA 153
435 GANGTAGNCGCGGAGGCTG--GAGAAASATNG--GSGGAGGTGGDQ--GSGGAGGTGAGGT 489
154 AAAAGAGCGGCTGGGLGSGAGAGAGAGAGCGGCGGLGSGAGRGSGGQ--GAGAA 199
490 GGAAGDGGGCGGAGGAGGAGGAGAGGTGAGNGGNTGTAGTGAAGAGNGAAGKGA 549
200 GGLGSGAGCGGCGGGLGSGG-----AGRGGLGCGG-----A 230
550 GCGGCTGGCTGGCGGAGCGGAGCTGCDRTVGGCTVPASGGCGGAGGAGGCGGAG 609
231 GAAAAAGAGGCGG-----GCGAGAAAAAGAGAGCGGCGGLGSGAGRGQ--GAGA 281
610 GSGCGDGDAGTGGNGGNGNRNSGNTGTGAGGNGGGGAGAGAGAGCGGCTGNGGAG 669
282 AAAAAG-----GAGCGCTGGGQ--AGCGCTG--GLGSGAGRGGLGCGGAGAAAA 331
670 DAGDAGNGNGNGTGGNGGNGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
332 GAGAGGGLGCGGAGAAAAAG-----AGCGGGLGCGGAGAGAGAGAGAGAGAG 380
730 TSSGCGAG 789
381 GVRGGYGGGLGSGAGRGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 425
790 G--DGGNGGDDGFGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
426 GAGGGLGCGGAGAAAAAG-----GCGCTGGG-----GAGAGAGAGAGAGAG 462

Db 848 GTGNGGTGTGTGAG 899
RESULT 8
ID 044358 PRELIMINARY; PRT: 871 AA.
AC 044358;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN FLAG.
OS Nephila clavipes (Orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RX MEDLINE=98153262; PubMed=9480768;
RA Hayashi C.Y., Lewis R.V.;
RT "Evidence from flagelliform silk cDNA for the structural basis of
RT elasticity and modular nature of spider silks."
RL J. Mol. Biol. 275:773-784(1998).
DR EMBL: AF027972; AAC38846.1; -
DR InterPro: IPR000087; -
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
FT NON_TER 871
SQ SEQUENCE 871 AA; 71039 MW; 1FA1E3B7E0C5983A CRC64;

Query Match 34.1%; Score 921; DB 5; Length 871;
Best Local Similarity 41.0%; Pred. No. 5.7e-47;
Matches 245; Conservative 22; Mismatches 212; Indels 118; Gaps 22;

19 GGLGSGAG 71
245 GCGAGGAG 303
72 GCGYGGGLGSGAGRGGLGSGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 130
304 GCGAGGCGYGGAG 359
131 GY--GGLGSGAGRGSGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 174
360 GYGPAG 419
175 RGGLGSGAG 228
420 GAGLGGAGGAG 479
229 GAG 261
480 GAG 539
262 G-----GYGGLGSGAGRGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 304
540 GNNYGPGRSGGCGYGGGSGGCGYGGGSGGCGYGGGSGGCGYGGGSGGCGYGGG 529
305 GGLGSGAGRGGLGGO-----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 354
600 GGVGTGVPAG 659
355 -----GCGGLGGO-----GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
660 GCPYGGAG 717
398 GCGAG 455
718 GGTGPG--GAGPGGCTGPGGAG--GCTGPGGSGGAGAGAGAGAGAGAGAGAGAG 774
456 SAASAAASRLSPQASSRVSAVSNLVASGPTNSAALSTISNVVQIGASNDGLSG 512

```

Db 775 SESGYPGSGPGSGPBGSGPG---GSGPGYCGPGSGPSPSFVP--GSGPGGSG 826
RESULT 9
ID 053559 PRELIMINARY; PRT: 1489 AA.
AC 053559;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PGSS-FAMILY PROTEIN.
GN RV3514 OR MV023.21.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,
Hornby T., Jagers K., Kiroh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutler K., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; AL022022; CAAL1751.1; -.
DR Unclust; RV3514; -.
DR InterPro; IPR000084; -.
DR InterPro; IPR002173; -.
DR Pfam; PF00934; PE; 1.
DR PRODOM; PD001223; -.
DR PROSITE; PS00583; PKB_KINASES_1; UNKNOWN_2.
SQ SEQUENCE 1489 AA; 115754 MW; 6855CBA1C3CBAF3A CRC64;
Query Match 33.8%; Score 912.5; DB 2; Length 1489;
Best Local Similarity 39.1%; Pred. No. 2.9e-46;
Matches 263; Conservative 19; Mismatches 202; Indels 189; Gaps 34;
OY 2 ASMTGCGQMGRIIRIRYGVGLGSGAGQ--GAGAAAAAGAGGCGGGLGSGAGRGGO 59
Db 178 AGGAGGWLFG---VGGAGGAGGAGGCTGTGAGGPGGLTWGGGAGGAGGCTGTGAGGRA 234
OY 60 ---GAGAAAAA--GG---AGGCGYGLGSGQ---AGRGGLGSG--GAGAAAAAGCV 103
Db 235 ELRLGAGAGAGAGGCTGTGCGATGTCGHHGVGGDGMWLPAGAGAGAGGAGAGSDGA 294
OY 104 --GCGCGGAGAGGAGCA--AAAAAGAGGCGYGLGSGQ-----AGRGSGGCG- 148
Db 295 LGTGTGTGTGTGAGAGAGRGLLLGAGGCGGGLGAGCGGTGTGAGGDDVLTGAGGTGCGK 354
OY 149 -----GAGAAA-----AAAGGAGGCGYGLGSGAGRGGLGSGGAAAAAAAGAG 195
Db 335 GVGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 413
OY 196 -QGCGYGLGSGQ-----AGGCGYGLGSGQ-----GRGGLGSGQ-----AG 231
Db 414 FAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 473
OY 232 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 275
Db 474 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 533
OY 276 GQ--GAGAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 328
Db 534 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 592

```

```

OY 329 AAAGAGCGGLGSG--GAGAAAAAGAGAGGGLGSGQ-----AGGAGA 370
Db 593 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
OY 371 AAAAAAGAGV-----ROGTYGLGSGQA-----GRCGAGAAAAAAG 410
Db 653 GKAGSSSAGAGTSSSGAGTGTGTGAGAGAGADADPAGTGTGTGTGAGAGAGAG 712
OY 411 G-AGCGTGTG---GGCGV-----GAGGLGSG--GAGAAAAAGA----- 443
Db 713 GAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 772
OY 444 --GCGYGVGSSGASAAASRLSSPOASSRVSSAVSNLVASSPTNSALSTISNVVS 501
Db 773 IGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 810
OY 502 QIGASNPGLSGCD 514
Db 811 --GVGAGAGGCGGD 821
RESULT 10
ID 09NHM4 PRELIMINARY; PRT: 2249 AA.
AC 09NHM4;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE FLAGELLIFORM SILK PROTEIN (FRAGMENT).
GN FLAG.
OS Nephila clavipes (orb spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Araneoidae; Tetragnathidae; Nephila.
OX NCBI_TaxID=6915;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20156766; PubMed=10688794;
RX Hayashi C.Y., Lewis R.V.;
RT "Molecular architecture and evolution of a modular spider silk protein
gene."
RL Science 287:1477-1479(2000).
DR EMBL; AF218621; AAF36090.1; -.
DR InterPro; IPR000087; -.
DR InterPro; IPR002029; -.
DR PROSITE; PS00138; SUBRTILASE_SER; UNKNOWN_1.
FT NON_TER 1
FT TER 2249
SQ SEQUENCE 2249 AA; 174871 MW; 88C8B168A147CDA0 CRC64;
Query Match 33.3%; Score 900; DB 5; Length 2249;
Best Local Similarity 36.0%; Pred. No. 2.2e-45;
Matches 245; Conservative 27; Mismatches 220; Indels 188; Gaps 24;
OY 6 GCGMGRIIRIRY--GSLGAGAGAGAGAAAAAGAGGCGY-----GGLSGGAGR 56
Db 224 GGVGPGAGPGYGPAGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGPGAGAGL 283
OY 57 G-----GAGAAAAAAGAGAGGCGY--GGLSGGAGRGGL----- 89
Db 284 GAGAPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
OY 90 GGGAGAAAAAGV--GCGGLGSGAGQ-----GAGAAAAA 123
Db 344 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 403
OY 124 AGAGAGQ-----GYGLGSGAGAGGSGGAGAGAAAAAGAGAGGCGY-----GGLG 169
Db 404 PGAGPGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 463
OY 170 SGAAGRGGLGQ-----GAGAAAAAAGS-----AGGCGYGLGSGQA----- 207

```


	Query Match	32.5%	Score 877.5;	DB 5;	Length 1468;	
	Best Local Similarity	36.1%;	Pred. No. 3.2e-44;			
	Matches 262;	Conservative 51;	Mismatches 182;	Indels 231;	Gaps 28;	
OY	2 ASMTGCGQMGRTRINGYGGLGGAGCAGCGAATAAAAGGACGGCGGYGCLGSOG-----AGR 56	:::::	: : :	: : :	:	
Dd	357 SASVTSAGAGLGVGAAGAASGLGGL---GAGASAAGSA-GAGJAGGIGAGSGSSSNAASAAS 412					
OY	57 GGCGAG-----AAAAAGCAAGCG-----YGCLSQG-----83		: :			
Dd	413 GASGAGEVIYIDRRSSAAAASSGASGLGLGLGLGPYGTICLTNCTVSASALGAGLG 472					
OY	84 -----AGRGGLGCAGC-----AAAAGCVGGGLGCAGC-----CG 116					
Dd	473 GVGTAGAGLGLGLGAGAYAVCPACAGLGGVAGCGSSGSSAASARSAPAVIIEDG 532					
OY	117 AGAAAAAGGAGCGGYGGL-----GSGGAGRGGSGGCGA 150					
Dd	533 SSMASMAAAGSCASGLGGLGCMKPLGIGTENEYSNASATCSAAGSTGAGLGGSGAMGS 592					
OY	151 GAAAAAAG-----GAGCGGYGGL-----168					
Dd	593 SMSASASGMAPAVIYIEDGSSAASAAAGSGASGLGLGCMKTLCGTCIPNGVSSAS 652					
OY	169 -----GSGGARGLGGCG-----AGA-----AAAAAGACGGCYG 201					
Dd	653 ATGSAAGSTGSLGCGSAAAGSLASASCAMAAPVYIEDSSAASAAAASGASGVGG 712					
OY	202 LCGGAGCGCGYGCGLGSOGARCGRLGCAGGAAAAAAGACGGCGLGGCG---AGAAAAAGG 258					
Dd	713 L-GLCA-LGPLGITPTPGASMSAGAGLGVGAAGTSLGGLTGIVGASTAGSAGNLGG 770					
OY	259 ACGGCGYGLGSGOAGRCGOAG-----AAAAAGCACGGYGCAGC-QGYYGGL 307					
Dd	771 ICAGGSSGSSNAASAGSAGCEVIYIDRRSSAASAAAGSGASGPGGLGLGWBPGLGI 830					
OY	308 GSQGARGGLGCGCGAANAAGCAGCGGCGAGCAAAAAGCAGCGGCGCGAGCG 367					
Dd	831 GPICGSSASASACAGLG---GYCAMOTSGIGGIG--GAGSAAGSAMGAGIGGIGAGSGSSGGS 885					
OY	368 AGAAAA-----AAAAAAGVRCCGCGGLGSGOAG-RCCGCA-GA AAA 408					
Dd	886 SVASAAAGSTSGAGEVIYIDRRSSAASAAAAGSGASGLGLGLGLPYGTICTLNVSAS 945					
OY	409 AGCAGCGGTGCLGCGCGVAGAGLGCAGAAAAGCAGCGGCGVGS-GASMAAASRLSS 467					
Dd	946 ALGAGLGVGTAGACAGL--GGLGCTGASAMGAGAGLGVGAGCGFGSSAASAGASGA 1003					
OY	468 PQA-----SRVSVSNLVASGP-----TNSALSRTISNVSOIQAS 506					
Dd	1004 GEVIYIDRRSSAASAAAGSGAGCGGLGLGWYPLGIGTPNGVSSASLTGSAAGSTGA- 1062					
OY	507 NPGLSG 512					
Dd	1063 ---GLGG 1066					
RESULT	15					

	Query Match	32.2%	Score 868.5	DB 2:	Length 1660;	
	Best Local Similarity	41.5%	Pred. No. 1,2e+43;			
	Matches 245;	Conservative	22;	Mismatches 175;	Indels 149;	Gaps 30.
Qy	17 GYGGLGOGAGOCAGAAAAAAGACGCQGYGGLGSQ----	-GAGRGOAGAAAAAAGAG	72			
	:	:				
Dd	649 GAGGAGGAATGTGATGATYHSHSGNGKGKGNADATYAANCGKAGAGNGELTGDCG		708			
Qy	73 QGGYGGLSQGA-----	-GRGLGCOGAGAAAAGS----	102			
	:	:				
Dd	709 AGCDGSGSAANGANGAVNVEDGADGTLSCPGEGSEANGCGGGVGGGAGAGAGDGGAGSS		768			
Qy	103 -VGQGLLGGQG--AGQCAGA-AAAAAGCA-----	-GQGGYGLLSQGA--GRGCSG	146			
	:	:				
Dd	769 AAGSGGNGRGRBAGQAQAGGAGAGAGAGSVSGDGPBGAKGAGAGAGAGAGAGGCGGCGK		828			
Qy	147 GCGAGAAAAAACGACGCGYGLSLDSOGRGLGQCGAGAAAAAAGACGCQGYGGLGCGQ		206			
	:	:				
Dd	829 AAGABSAEAVCGAGCKGGBGQYGVIV--GQGGTGGDS--GAGGAAPAQVSHVGVGVGBG		886			
Qy	207 ----AGQGGYGLLSQGA-----	-GRGLGQ----QAGAAAAAG-----GACQCG	243			
	:	:				
Dd	887 GLGAGAGNGDGHDGSDGDGDGCPGAGGLGLGLGSDGNGTTPAAASVDASDHGPGSSG		946			
Qy	244 LGGQGA-GAAAAAAGGA-QQGGYGC-----LGSQGA-GRGGQCGAAAAAAGGA-----		289			
	:	:				
Dd	947 NGNGNGNGAQASVAAGGAGGNGDDGAGNRVYGGGAGCNGGDDAAGCANNGANSAPGSDALA		1006			
Qy	290 -GQ--GGYGAQG-----AGQGGYGLLSQGARGRLGQ--GAGAAAAAAGAGCGQG		338			
	:	:				
Dd	1007 LGQPGNGGQDAGAGAGAGAGAGAGAGSVSGDGGAGNGAGNGAGNGCVGASGGAGAG		1066			
Qy	LGG----QAGAAAAAAGAGAGCGGLGQDAGQACAGAAAAAAGAAAA---GCVROGYGTG		391			
	:	:				

```

Db 1067 ANGIDSTGTGGAGGGGGGAGGVGHHGDPGVGAAPSGTVGSHGTGV--GGDGLG 1124
QY 392 SQG--AGRGQ-----GAGAAAAAGAGGQGTGLG-----OGV----- 425
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1125 GAGGVGAGGNGGIGITYCGAGAGNGGDPGAGRGGLGDSGNGTSAANGVDASKHGP 1184
QY 426 ---GAGLGGGAGAAAAVG-AGGGYGG-----VSSGASASASAAA 462
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1185 LTGGDGVGNGAKAAAAAGDGGGCGDGNAGLFGDGGAGGDADGTAAEA 1235

```

Search completed: July 3, 2001, 15:05:41
 Job time: 606 sec